



Article Deep Learning-Based Approaches for Enhanced Diagnosis and Comprehensive Understanding of Carpal Tunnel Syndrome

Marwa Elseddik ^{1,2}, Khaled Alnowaiser ^{3,*}, Reham R. Mostafa ^{2,4}, Ahmed Elashry ⁵, Nora El-Rashidy ^{6,*}, Shimaa Elgamal ⁷, Ahmed Aboelfetouh ^{2,8} and Hazem El-Bakry ²

- ¹ Department of the Robotics and Internet Machines, Faculty of Artificial Intelligence, Kafrelsheikh University, Kafrelsheikh 33516, Egypt; marwaelsadeek@ai.kfs.edu.eg
- ² Department of Information Systems, Faculty of Computers and Information, Mansoura University, Mansoura 35516, Egypt; reham_2006@mans.edu.eg (R.R.M.); elfetouh@mans.edu.eg (A.A.); elbakry@mans.edu.eg (H.E.-B.)
- ³ College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al Kharj 11942, Saudi Arabia
- ⁴ Research Institute of Sciences and Engineering (RISE), University of Sharjah, Sharjah 27272, United Arab Emirates
- ⁵ Department of Information Systems, Faculty of Computers and Information, Kafrelsheikh University, Kafrelsheikh 33516, Egypt; ahmed_elashry@fci.kfs.edu.eg
- ⁶ Department of Machine Learning and Information Retrieval, Faculty of Artificial Intelligence, Kafrelsheikh University, Kafrelsheikh 33516, Egypt
- ⁷ Department of Neuropsychiatry, Faculty of Medicine, Kafrelsheikh University, Kafrelsheikh 33516, Egypt; shimaa_elgamal@med.kfs.edu.eg
- ⁸ Delta Higher Institute for Management and Accounting Information Systems, Mansoura 35511, Egypt
- * Correspondence: k.alnowaiser@psau.edu.sa (K.A.); noura.alrashidy@ai.kfs.edu.eg (N.E.-R.)

Abstract: Carpal tunnel syndrome (CTS) is a prevalent medical condition resulting from compression of the median nerve in the hand, often caused by overuse or age-related factors. In this study, a total of 160 patients participated, including 80 individuals with CTS presenting varying levels of severity across different age groups. Numerous studies have explored the use of machine learning (ML) and deep learning (DL) techniques for CTS diagnosis. However, further research is required to fully leverage the potential of artificial intelligence (AI) technology in CTS diagnosis, addressing the challenges and limitations highlighted in the existing literature. In our work, we propose a novel approach for CTS diagnosis, prediction, and monitoring disease progression. The proposed framework consists of three main layers. Firstly, we employ three distinct DL models for CTS diagnosis. Through our experiments, the proposed approach demonstrates superior performance across multiple evaluation metrics, with an accuracy of 0.969%, precision of 0.982%, and recall of 0.963%. The second layer focuses on predicting the cross-sectional area (CSA) at 1, 3, and 6 months using ML models, aiming to forecast disease progression during therapy. The best-performing model achieves an accuracy of 0.9522, an R2 score of 0.667, a mean absolute error (MAE) of 0.0132, and a median squared error (MdSE) of 0.0639. The highest predictive performance is observed after 6 months. The third layer concentrates on assessing significant changes in the patients' health status through statistical tests, including significance tests, the Kruskal-Wallis test, and a two-way ANOVA test. These tests aim to determine the effect of injections on CTS treatment. The results reveal a highly significant reduction in symptoms, as evidenced by scores from the Symptom Severity Scale and Functional Status Scale, as well as a decrease in CSA after 1, 3, and 6 months following the injection. SHAP is then utilized to provide an understandable explanation of the final prediction. Overall, our study presents a comprehensive approach for CTS diagnosis, prediction, and monitoring, showcasing promising results in terms of accuracy, precision, and recall for CTS diagnosis, as well as effective prediction of disease progression and evaluation of treatment effectiveness through statistical analysis.



Citation: Elseddik, M.; Alnowaiser, K.; Mostafa, R.R.; Elashry, A.; El-Rashidy, N.; Elgamal, S.; Aboelfetouh, A.; El-Bakry, H. Deep Learning-Based Approaches for Enhanced Diagnosis and Comprehensive Understanding of Carpal Tunnel Syndrome. *Diagnostics* **2023**, *13*, 3211. https://doi.org/ 10.3390/diagnostics13203211

Academic Editor: Jae-Ho Han

Received: 31 August 2023 Revised: 3 October 2023 Accepted: 5 October 2023 Published: 14 October 2023



Copyright: © 2023 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/). **Keywords:** carpal tunnel syndrome (CTS); deep learning (DL); machine learning (ML); Adam optimizer; cross-sectional area (CSA); statistical analysis

1. Introduction

1.1. Overview

Carpal tunnel syndrome (CTS) is a prevalent type of compressive mononeuropathy that is due to the entrapment of a nerve. Studies indicate that nearly 90% of all cases of entrapment neuropathy led to CTS. Potential contributing factors to CTS include the presence of digital flexor tendons, wrist bone and the transverse carpal ligament, as well as oedema, strenuous manual activity, hormonal changes, and tendon inflammation. In severe cases, weakness in the hand may result from injury to the motor fibers of the median nerve (MN). The exact cause of CTS remains uncertain, but MN compression, biochemical changes, oedema, and tissue adhesion surrounding the MN are commonly considered plausible explanations. The therapy recommendations for CTS vary depending on the severity of the condition, which ranges from a conservative approach for mild and moderate cases to surgical surgery for severe cases. Conservative therapy may be beneficial for most cases with mild to moderate CTS; however, a Cochrane Review concluded that such treatments had only short-term or limited effectiveness in severe cases. Surgical decompression is considered the main solution advocated for severe CTS or patients who have an unsatisfactory response to conservative treatment. As a result, innovative intervention during the presurgical stages of CTS is required [1,2].

The carpal tunnel, located at the base of the palm, is formed by the eight carpal bones and the transverse carpal ligament (TCL). It accommodates several structures, including eight digital flexor tendons, the flexor pollicis longus tendon, their flexor synovial sheaths, and the median nerve (MN). Compression of the median nerve can occur if there is an increase in the volume of these structures, leading to nerve ischemia and resulting in pain and paresthesia. Symptoms of carpal tunnel syndrome (CTS) primarily affect the lateral three fingers and the lateral half of the ring finger, while the palm remains unaffected due to the sensory cutaneous branch of the median nerve being unaffected by the pressure changes within the carpal tunnel [3,4].

1.2. Problem Statement

The relationship between CTS and artificial intelligence (AI) is progressing rapidly, especially in the field of medical diagnosis and treatment. Deep learning (DL) algorithms can be trained on large datasets of patient information, including medical history, symptoms and diagnostic test results, to identify patterns and features that may be difficult for human clinicians to detect [5]. This way can lead to earlier and more accurate diagnoses of CTS, as well as the ability to predict the progression of the disease and create personalized treatment strategies for individual patients. AI can analyze various types of data, including imaging data such as magnetic resonance imaging (MRI) or ultrasound images and electromyography (EMG) data [5,6]. Sensory and motor nerve conduction studies (NCS) are valuable tools in the diagnosis and staging of CTS. They provide objective measurements of nerve functionality; help differentiate CTS from other conditions and assist in determining the severity of the condition. By incorporating these studies into the diagnostic process, healthcare professionals can make informed decisions regarding the treatment and management of CTS patients [7]. DL models can identify changes in the MN or other structures that may indicate CTS, which leads to earlier diagnosis and treatment and improves the accuracy and efficiency of diagnosis. The relationship between CTS and AI is in its early stages but has significant potential to improve patient outcomes and advance the field of medical diagnosis and treatment. Recent studies have ignored the effect of clinical, personal and historical data on disease diagnosis and treatment [8,9]. This study addresses these limitations by using a DL model to diagnose CTS based on a

combination of patient history, personal data, clinical examination data, NCS and CSA from ultrasound images.

1.3. Study Objectives

The objectives of this study are as follows:

- i. Utilize AI techniques to develop a model that support help medical experts in distinguish between CTS patients and nonpatients effectively and efficiently.
- Explore the role of patient data from the Boston Carpal Tunnel Questionnaire (BCTQ), NCS and CSA from ultrasound images in the CTS diagnosis process and treatment monitoring.
- iii. Develop AI model for supporting medical experts in the treatment process by predicting the cross-sectional area (CSA) of the MN after 1, 3 and 6 months of hydro dissection injection to determine the effectiveness of the injection treatment in improving patient outcomes.

1.4. Paper Organization

The rest of the paper is organized as follows. Section 2 provides a comprehensive literature review. Section 3 outlines the methodology used in the study. Section 4 focuses on the clinical diagnosis of CTS. Section 5 describes the dataset utilized in the study. Section 6 presents the proposed work in detail. Section 7 presents the results and discussion. Section 8 compares our study with other relevant works. Section 9 presents the model explanation. Section 10 concludes the paper and highlights areas for future research.

2. Related Work

CTS is a common condition that affects the hand and wrist. It occurs when the MN, which runs through the carpal tunnel, becomes compressed or squeezed. CTS presents various clinical manifestations, including pain, numbness, tingling, weakness, swelling and sensory changes in the hand and fingers. Diagnostic studies for CTS include NCS and ultrasonography. NCS is a standard diagnostic test that measures the speed and strength of electrical signals along the MN. It helps confirm the presence of nerve damage and assess the severity of CTS. In the meantime, ultrasonography is a non-invasive imaging technique that provides detailed images of the carpal tunnel and surrounding structures. It is useful in identifying structural abnormalities, such as thickening of the transverse carpal ligament or swelling or cysts. These diagnostic tools aid in accurately diagnosing CTS and determining the appropriate course of treatment. Regarding treatments, nonsurgical options include conservative management with wrist splinting, activity modification, physical therapy, and oral medications to alleviate pain and inflammation. In cases where traditional measures are ineffective, corticosteroid injections may be administered. Surgical intervention, such as carpal tunnel release (CTR) surgery, may be considered a last resort to relieve compression on the MN. AI algorithms have shown promise in diagnosing CTS by analyzing various data sources, including electrodiagnostic tests, imaging studies and patient-reported outcomes. These algorithms can aid in identifying CTS with precision and accuracy. AI can assist in developing personalized treatment plans for individuals with CTS by analyzing patient data, medical history, and treatment outcomes. It can help optimize treatment approaches and improve patient outcomes. AI algorithms can be utilized to evaluate the effectiveness of CTS treatments by analyzing changes in symptoms, functional outcomes, and patient-reported data. This way can provide valuable insights into treatment response and guide further interventions if needed. Furthermore, statistical tests such as ANOVA and t tests have been used to evaluate whether patient health status changes throughout the treatment process. These tests help identify the presence of CTS with precision. Several relevant works suggest models for diagnosing CTS, as illustrated in Table 1.

Authors	Method	No. Cases	Evaluation Measures (%)
Hoogendam et al. [10]	gradient boosting machines	2119 patients	AUC: 0.7820
Park et al. [11]	XGB	1037 patients	Accuracy: 76.6
Tsamis et al. [12]	SVM	38 patients	Accuracy: 0.9513
Harrison et al. [13]	QuickDASH	1916 patients	Accuracy: 0.72
Ciobanu et al. [14]	Boston-CTS	53 patients	Sensitivity: 89.7
Smerilli et al. [15]	Mask R-CNN	103 patients 246 images	Precision: 0.86
Cosmo et al. [16]	Mask R-CNN	53 patients 151 images	DSC = 0.93
Shinohara et al. [17]	Efficient Net	100 patients, 10,000 images	Accuracy: 0.96
Wang et al. [18]	MNT-DeepSL	100 cases, 84 patients	Accuracy: 0.9
Faeghi et al. [19]	SVM	228 wrists from 65 patients and 57 controls	Accuracy: 90.1
Hafane et al. [20]	Localisation + PGVF	ultrasound images elicited from 10 videos.	DSC = 0.85

Table 1. The state of the art of CTS diagnosing.

Some diagnoses are based on numerical data. For example, Hoogendam et al. [10] conducted a study to develop a prediction model for assessing the probability of improvement of symptoms reported by patients after 6 months. The proposed results showed that the gradient-boosting machines surpassed the logistic regression (LR) and random forest (RF) models in predicting clinically relevant improvements in symptoms. The highest model had a sensitivity of 0.84 and a specificity of 0.55. However, the limitations of this study include the existence of missing data, which affects model performance. Park et al. [11] developed machine learning (ML) models such as RF and extreme gradient boosting (XGB) to classify the severity of CTS using clinical and electrophysiological features. XGB showed the highest accuracy in multiclass classification with a test prediction accuracy of 76.6%. Tsamis et al. [12] used five ML classifiers, namely, LR, support vector machines (SVMs), k-nearest neighbours, decision trees and Naïve-Bayes, based on conventional electrodiagnostic criteria in the clinical practice of CTS. The classification was verified through neurophysiological and clinical diagnoses. The highest accuracy of 0.9513 was achieved by the SVM classifier. The results demonstrate the potential for CTS identification, which can eliminate human errors in decision making. Harrison et al. [13] developed an ML model using QuickDASH to perform patient-reported outcome measures and clinical data. The algorithm that made the most accurate prediction of functional and symptomatic improvement had respective accuracies of 0.72 and 0.76. Ciobanu et al. [14] investigated two questionnaires, namely, Boston-CTS and six-item CTS. The Boston CTS questionnaire had higher sensitivity (89.7%) and positive predictive value (88.9%) than the six-item CTS questionnaire (76.9% and 75.0%, respectively).

Other diagnoses are based on ultrasound images. For example, Smerilli et al. [15] developed a CNN model called Mask R-CNN to predict the measurement of the MN using ultrasound images obtained at the level of the carpal tunnel. The CNN was tested on a dataset of ultrasound images from patients with CTS and showed promising results in terms of accuracy = 0.86, sensitivity = 0.88 and specificity = 0.86. Cosmo et al. [16] developed a CNN model called Mask R-CNN, which achieved a DSC of 0.93. Shinohara et al. [17] investigated the accuracy of three different DL models (Squeeze Net, MobileNet_v2 and EfficientNet) to predict CTS from ultrasound images of the MN. The highest model had an accuracy of 0.96, a recall of 0.94 and a precision of 0.99. Wang et al. [18] used deep

similarity learning that included preprocessing, feature extraction, similarity learning and nerve following. The approach achieved high accuracy in following the MN at 0.9. Faeghi et al. [19] developed an approach for diagnosing CTS using radiomics features extracted from ultrasound images. This approach was then compared with radiologists' assessment of CTS diagnosis. The study concluded that the automated approach achieved high accuracy in the diagnosis of CTS and outperformed radiologists in certain aspects. Hafane et al. [20] developed a CNN model. CNN is used to identify the region of interest around the nerve. The results of this study showed a median DSC of 0.85.

In this research, we used numerical datasets for several reasons. Numerical datasets are often easier and faster to analyze than ultrasound images. Processing and analyzing numerical data require less computational power and time, which allows for quicker and more efficient diagnosis and monitoring of CTS. Obtaining numerical datasets for CTS is also generally more accessible and cost effective than acquiring ultrasound images. Ultrasound imaging requires specialized equipment and expertise, whereas numerical data can be collected using simpler and more widely available tools, such as EMG or NCS. Despite the promising performance from most literature, several limitations should be addressed as follows:

- (1) Several CTS diagnosis studies have been conducted based on small and limited datasets (i.e., data aggregated from medical questionnaires), which may exclude important features (e.g., clinical examination, clinical history, and demographics).
- (2) Aggregating data according to specific conditions (i.e., women older than 40 years) limits the generalization ability of the developed model.
- (3) Most studies ignored the overlap between CTS and other diseases, which may affect model accuracy.

3. Methods

3.1. Deep Learning

Artificial neural networks are used to model and resolve complicated issues in the deep neural network (DNN). DNN processes input data and produces output predictions using several layers of interconnected nodes or neurons. A DNN typically has three layers: an input layer, hidden layer, and an output layer [21]. The input layer, which is the first layer. Every node in this layer receives an input, processes it, and then sends its output as the input to every node in the following layer. The final layer in a deep neural network generates the output prediction based on the learned features and weights, and there are typically no connections between nodes in the same layer. The layers between the input and output layers are referred to as the hidden layers in the middle section. Every node in each one carries out mathematical operations on the incoming data to produce output values, which are then sent on to the following layer. DNNs can have numerous hidden layers with various numbers of nodes in each layer. DL hidden layers can include multiple types, such as: (i) Dense Layers: This layer is a type of fully connected layer in which each node in the layer is connected to every node in the previous layer. The nodes in a dense layer perform a linear transformation on the input data and apply an activation function to generate an output value. This layer is used for feature extraction and classification tasks. (ii) Batch normalization: This layer normalizes the previous layer's output and applies a scaling and shifting operation to improve the stability and speed of training. (iii) Dropout: This layer randomly drops out a fraction of the neurons in the previous layer during training, which helps to prevent overfitting [5,22]. For multi-class classification problems, neural networks frequently use the SoftMax activation function. It creates a probability distribution over classes from a vector of input values. The result is a probability distribution that sums to one after each member of the input vector has been subjected to the exponential function and is normalized. To determine the most likely class for a given input, this function is frequently employed in the output layer of a neural network for multi-class classification problems. The probabilistic interpretation of the output that SoftMax offers is one of its benefits, but it can also be vulnerable to noise and outliers in the input data [23].

The SoftMax function can be mathematically expressed as follows:

$$P(y = j|x) = \frac{e^{x_j}}{\sum_{i=1}^{K} e^{x_i}}$$

$$\tag{1}$$

where P(y = j|x) is the probability that the input *x* belongs to class *j*, *x* is the vector of logits (input values), x_j is the *j*-th element of the input vector *x* and *K* is the total number of classes.

This study found that the most effective DL optimization algorithm was Adam (Adaptive Moment Estimation). Adam is a commonly used optimization algorithm in DL that minimizes the loss function during training. It is an extension of stochastic gradient descent (SGD) that integrates concepts from momentum and adaptive learning rates. Adam optimizes the parameters of the neural network by maintaining a running estimate of the first and second moments of the gradients. The first moment is the mean of the gradients, while the second moment is the uncentered variance of the gradients. Exponential moving averages are used to calculate these estimates, with recent gradients given more weight [23,24].

The Adam optimizer updates the parameters of the neural network using the following equations:

$$m_t = \beta * m_{t-1} + (1 - \beta) \times g_t \tag{2}$$

$$v_t = \beta_2 * v_{t-1} + (1 - \beta_2) \times (g_t)^2$$
(3)

$$\theta_{t+1} = \theta_t - \frac{(\alpha \times m_t)}{\sqrt{v_t + \epsilon}} \tag{4}$$

where m_t and v_t are the first and second moment estimates of the gradients at time step t, g_t is the gradient at time step t, thetat is the parameter vector at time step t, alpha is the learning rate, beta1 and beta2 are the exponential decay rates for the first and second moments, and epsilon is a small constant added for numerical stability.

3.2. Statistical Tests

The data collected in the study were analyzed using the SPSS version 23 for Windows[®] (IBM SPSS Inc., Chicago, IL, USA). SPSS provides a range of tools and techniques for data manipulation, descriptive statistics, inferential statistics, data visualization, and reporting such as:

3.2.1. ANOVA Test

ANOVA is a statistical process for comparing the means of various samples. It is like extending the *t*-test for two independent samples to more than two groups. The goal is to test for substantial variations across classes by analyzing the variances [25]. The hypothesis in the ANOVA test is comparing two independent estimates of the population variance. It is one of the most beneficial tests for disclosing significant information, especially when interpreting experimental results and identifying the impact of some elements on other processing parameters [26]. It assesses whether a statistical process produces useful results. It essentially allows you to choose whether to reject or accept a null hypothesis. Two factors are utilized to determine this in a two-way ANOVA test. A two-way ANOVA test makes the following assumptions: Firstly, the two testing variables should be independent. Secondly, the total variance should be homogeneous (volatility around the mean should be consistent). Finally, the variables should have a normal distribution. Assume that there are two populations: y_{11} , y_{12} , y_{13} , ..., y_{1n} and y_{21} , y_{22} , y_{23} , ..., y_{2n} . We have independent variables y_{ij} , $i = 1, 2, 3, \ldots, k$ and $j = 1, 2, 3, \ldots, n$, with mean μ_i and standard deviation of ∂ . In this test, we are mainly concerned with the null hypothesis.

The against the hypothesis is:

$$H_0: \mu_1 = \mu_{2=} \mu_K \tag{5}$$

y' refers to the grand mean, the mean of all the data points.

$$y' = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}$$
(6)

 s_i^2 represents the sample of the variance.

$$s_{i}^{2} = \frac{1}{n_{i} - 1} \sum_{j=1}^{n_{i}} \left(y_{ij} - y_{i}^{'} \right)^{2}$$
(7)

 $s_i^2 = MSE$ estimates the σ^2 . ANOVA is mainly centered on the idea of comparing the variations between two groups as well as the variations within samples.

3.2.2. Level of Significance

In this study, the *p*-value is used to measure significance levels. The *p*-value is also known as the probability value; it indicates how likely our results would have occurred assuming that the null hypothesis is correct [27]. This is accomplished by computing the likelihood of the test statistic, which is the number determined by a statistical test based on the data [28,29]. The degree of significance was assessed for all the above-mentioned tests. Results can be described as follows: nonsignificant if the *p*-value is higher than 0.05 (p > 0.05), significant if the *p*-value is lower than 0.05 (p < 0.05), and highly significant if the *p*-value is less than 0.001.

3.2.3. Kruskal-Wallis Test

The Kruskal-Wallis test is a non-parametric statistical test used to compare the effect of three or more groups on a continuous variable when its distribution is not normal in one or more groups. It is used to determine whether there are significant differences between the groups based on the ranks of the data rather than the actual values [30].

4. CTS Clinical Diagnosis

4.1. CTS Symptoms

Identifying appropriate symptoms is essential for diagnosing the presence or absence of CTS. These symptoms typically include numbness, tingling or a burning sensation in the volar areas, especially at night or after strenuous activities. Nocturnal symptoms are common amongst the majority of patients and may involve the entire hand or be limited to the thumb or the first two or three fingers. Patients with CTS often report a unique sensation of swelling in their hands, despite the absence of visible oedema. In some cases, NCS may reveal thenar atrophy and denervation. Other CTS symptoms may include writer's cramp, forearm pain, shoulder discomfort, cold sensitivity in the fingers or numbness in the third finger alone [31].

4.2. Clinical Examination

In the classic presentation of CTS, symptoms typically affect two or more of the first three fingers. Pain that extends beyond the wrist and involves the fourth and fifth fingers, along with wrist pain, may also be experienced. However, involvement of the palm or dorsum of the hand is not typically associated with CTS symptoms [32].

4.3. Motor Examination

Thenar atrophy is a late-onset condition that results in severe functional loss. Finger weakness combined with the inability to pinch or repeated dropping of gripped items implies the involvement of motor components. The loss of feeling up to a pinprick in the MN distribution frequently occurs before thenar atrophy. Thenar atrophy is rarely noticed by patients and may be unnoticed when evaluated by gazing down at the palm. However, it is easily discernible by comparing both palms together. In a study conducted by Phalen [33], atrophy of the abductor pollicis brevis, opponents pollicis and flexor pollicis brevis muscles was observed in 41% of hands. Amongst these muscles, the abductor pollicis brevis is the most commonly affected, and its function can be assessed using the 'pen test' [33], which can be a useful tool in diagnosing CTS.

4.4. Scoring System

The BCTQ is a patient-reported questionnaire that assesses symptom severity and the overall functional condition of CTS cases [34]. The questionnaire consists of two scales, the Symptom Severity Scale (SSS) and the Functional Status Scale (FSS), which assess the severity of symptoms and the degree of difficulty in performing daily tasks, respectively. The SSS contains 11 questions scored on a Likert scale of 1 to 5, whilst the FSS consists of eight questions scored on a scale of 1 to 5, where a score of 1 represents no difficulty and 5 indicates severe difficulty.

5. Dataset Description and Preparation

5.1. Data Description

5.1.1. Dataset Collection

The dataset was collected retrospectively from the Neurology Department at Kafrelsheikh University Hospital, Egypt, between April 2019 and April 2020. It included 160 patients who were divided into two groups: those diagnosed with CTS and those with similar symptoms. The study was submitted for IRB approval (Faculty of Medicine, Kafrelsheikh University), and patients' confidentiality and privacy were ensured throughout the study [35].

5.1.2. Study Cohorts

The dataset for the study on CTS patients was collected based on specific inclusion criteria: (i) participants aged 20 to 60 years; (ii) manifestations of CTS; (iii) NCS showed delayed sensory or motor conduction of the MN; (vi) patients who did not respond to medical treatment after at least 3 months of symptom onset were included, and pregnant women were excluded from the study.

5.1.3. Data Aggregated for Each Patient

Personal and historical data: The patients' historical and personal data will be obtained and recorded, including information such as age, gender, body mass index (BMI), occupation, marital status, lifestyle habits and any relevant family history of similar conditions. Any prior medical or surgical problems will be noted as well.

Medical questionnaire: A computerized CTS sheet, including all variables of the BCTQ, was used to review all patients. The BCTQ is a reliable and valid tool used to evaluate the severity of symptoms and overall patient function. It includes two models, namely, SSS and FSS, which can be used together or separately.

Ultrasonographic examination: A single sonographer obtains the CSA using the tracing feature on the US machine (in mm^2 at the distal wrist crease) without weaving between each fascicle [36]. This method is more accurate than the ellipsoid approach. CTS is categorized by an MN area of >9 mm². According to the classification of El Miedany et al. [37], we classified the CTS severity scale based on CSA as follows: mild if CSA is up to 13.0 mm², moderate if CSA is between 13.0 and 15.0 mm² and severe if CSA is more than 15.0 mm² [38]. Three measurements were taken, and the average value was used for statistical analysis. All patients were administered injectate consisting of 1 mL lidocaine, 2 mL (8 mg) dexamethasone and 2 mL normal saline containing 300 IU hyaluronidase. We compared the CSA of the CTS cases with another 80 non-CTS volunteers who exhibited similar symptoms from the Neurology Department, Kafrelsheikh University Hospital

inpatient and outpatient clinics after matching for age and sex. Finally, Nerve Conduction Studies (NCS) were performed on all patients included in our study.

5.2. Dataset Preparation

5.2.1. Outlier Detection

Outlier detection involves identifying abnormal items among normal ones. It is a crucial step in data preparation as it can affect the performance of clustering and classification models. Different statistical techniques are used to address the issue, including proximity-based and distance-based methods. While these methods are effective, in this study, we relied on the expertise of a medical professional to identify and handle any data outliers.

5.2.2. Data Imputation

Missing values are pervasive in medical data due to corruption or collection errors. They can adversely impact the performance of classifiers by introducing bias. Various methods exist for filling in missing values, including basic techniques like mean, max, min, and the most frequent item. In our study, we encountered a small number of missing values in each column, ranging from 2 to 5 [39]. To achieve high accuracy in the imputed data, we employed a variable strategy known as multivariate imputation by chained equations (MICE) [40,41].

5.2.3. Data Scaling

Data scaling methods in machine learning are employed to address the importance of scalability and ensure accurate outcomes while minimizing uncertainties, incorrect predictions, and additional costs or processing time. One common approach to data scaling involves transforming the minimum value of a feature to 0 and the maximum value to 1 [6,9,42–49]. In our study, we applied data scaling using the following equation:

>

$$\zeta' = \frac{\mathbf{x} - \overline{\mathbf{x}}}{\delta} \tag{8}$$

6. Proposed Work

The proposed model for identifying and predicting CTS diagnosis at the carpal tunnel is divided into four stages, as shown in Figure 1. The first stage involves aggregating the necessary data, which includes patient history, personal data, clinical examination data, CSA from ultrasound images, NCS and BCTQ. The second stage is the data preprocessing stage, which involves cleaning, formatting, and standardizing the data to prepare them for analysis. The third stage is utilizing AI models to predict CTS diagnosis and monitor progression during treatment. We built a classification model to predict CTS diagnosis. DNN was utilized to build efficient models in terms of several evaluation metrics. Then, we built an ML model to predict the CSA after 1, 2 and 6 months. Lastly, the fourth stage involves utilizing statistical tests (level of significance, Kruskal-Wallis test and two-way ANOVA test) to evaluate whether a significant change in patient health status occurs through the treatment process. The results of the statistical tests showed highly significant changes in patient scores, including SSS, FSS and CSA after 1, 3 and 6 months of postinjection. Overall, the proposed model provides a comprehensive approach to diagnosing CTS based on AI techniques and advancement on the medical and AI sides.



Figure 1. Proposed method architecture.

7. Results and Discussion

7.1. Evlaution Metrics

Various metrics are employed to assess the performance of the model classification model and regression model, encompassing the following measures for classification (accuracy, precsion, recall, f_meausre and area under the roc_curve. While other evaluation metrics utilized for regression include mean square error (MSE) Mean absolute error (MAE), median absolute error (MedAE) and the R2 score Table 2 presents a comprehensive overview of these metrics and the mathematical formulas used to calculate them.

Metric	Abbreviation	Equation
Accuracy	ACC	$\frac{tp+tn}{tn+fn+tn+fn}$
Precision	Р	$\frac{tn}{tn+fp}$
Recall	R	$\frac{tn}{tn}$
F1-score	F1	$\frac{2(P * R)}{P + R}$
Mean_absolute error	MAE	$\frac{1}{n}\sum_{i=1}^{n} y_i-\hat{y}_i $
Mean_square_error	MSE	$\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$
Mediam absolute error	MedAE	median $(y_1 - \hat{y}_1 , y_2 - \hat{y}_2 , \dots, y_n - \hat{y}_n)$
R ² _score	R^2	$1-rac{\sum_{i=1}^n \left(y_i-\widehat{y}_i ight)^2}{\sum_{i=1}^n \left(y_i-\overline{y} ight)^2}$

Table 2. Evalution metrics of the model.

7.2. Predicting CTS Diagnosis

In this section, we evaluate the performance of DL in predicting CTS diagnosis. Three different DL models are built with three different optimizers: gradient descent (GD), adaptive gradient algorithm (Adagrad) and Adam. Figure 2 clarifies the plot of the architecture of the DL model. It includes six layers, including an input layer that has 37 inputs, four hidden layers with the ReLU activation function and an output layer with the Adam optimizer and sigmoid activation function. Table 3 shows the hyperparameters of the DL model. First, we tried the proposed model without historical data to check the effect of the historical data on the overall performance of the model. Figure 3 shows the learning and model without historical data. From Table 4, we can observe the best performance was obtained (ACC = 0.829%, precision = 0.823%, R = 0.857%, F-measure = 0.846% and AUC = 0.837%).

Table 3. Hyperparameters of DL model.

DL Model Hyperparameters	Value
Input layer	37 unit
Number of layers	6
Regularization	L2 = 0.1
Dropout	0.1
Batch size	32
Activation function in hidden layers	ReLU
Number of epochs	10
Activation function in output layer	Sigmoid
Optimizer used	ADAM

Table 4. Result of Predicting CTS Diagnosis without historical data.

Model	Optimizer	Accuracy	Precision	Recall	F1-Score	AUC
Model 1	GD	0.805%	0.823%	0.821%	0.812%	0.812%
Model 2	Adgard	0.812%	0.832%	0.833%	0.832%	0.824%
Model 3	Adam	0.829%	0.823%	0.857%	0.846%	0.837%



Figure 2. Architecture of the DL model.



Figure 3. Evaluation of DL model (a) model accuracy (b) model loss without historical data.

Second, we explored the effect of all data including (historical, clinical data and medical score). From Table 5, we can observe the following: the lowest was obtained from GD (ACC = 0.935%, Precision = 0.953%, R = 0.944%, F-measure = 0.947% and AUC = 0.963%) and utilizing Adagard enhanced the performance by approximately 1–2% (ACC = 0.955%, Precision = 0.963, R = 0.946%, F-measure = 0.957% and AUC = 0.963%). The best performance was obtained from Adam optimization, and the model achieved the best performance in terms of several metrics (ACC = 0.969%, precision = 0.982, R = 0.963%, F-measure = 0.974% and AUC = 0.972%). Figure 4 shows the learning and model with historical data. The results confirm the significant effect of the historical data.

Model	Optimizer	Accuracy	Precision	Recall	F1-Score	AUC
Model 1	GD	0.935%	0.953%	0.944%	0.947%	0.963%
Model 2	Adgard	0.955%	0.963%	0.946%	0.957%	0.963%
Model 3	Adam	0.969%	0.982%	0.963%	0.976%	0.972%

Table 5. Result of Predicting CTS Diagnosis with historical data.



Figure 4. Evaluation of DL model (a) model accuracy (b) model loss with historical data.

7.3. Predicting CSA Progression for Patients (1 Month, 3 Months, and 6 Months)

In this section, we performed six experiments on three different datasets to predict the CSA after 1, 3 and 6 months based on the ML models RF and multilayer perceptron (MLP). The hyperparameters of the ML model are shown in Table 6, and the models evaluated in terms of several evaluation metrics (training score, testing score, R2 score, mean absolute error and median square error) are shown in Table 7. In the first experiments, we utilised all data aggregated in the first month after injection, including the initial CSA, FSS initial stage, SSS initial stage, handshaking, and sensory symptoms. Utilising RF achieved adequate performance of 0.863, 0.884 and 0.981 in terms of training score, testing score and R2 score, respectively. The results quietly improved using MLP. The second experiment uses all the data from the first experiment in addition to the CSA in the first month, FSS and SSS that was calculated in the third month. The additional data increase the model performance. The same is true for experiment three, which utilises all the precious data in experiments 1 and 2 in addition to some extra data aggregated after 6 months. The performance of prediction after 6 months improved more than the others. The best performance was obtained from MLP (ACC = 0.9522, R2 score = 0.667, MAE = 0.0132, MDSE = 0.0639). Figure 5a,b show the residuals and prediction for the MLP regressor model.

Model	Hyperparameters for Machine Learning Models
MLP RF	Activation = ReLU, Batch size = 32, Regularization = None, Number of epochs = 70 n_estimators = 100, max_depth = 5

Table 7. Results of regression model for pre	dicting CSA.

Table 6. Hyperparameters of ML model.

CSA Over Time	Algorithm	Accuracy (Train)	Accuracy (Test)	R2 Score in Train	R2 Score in Testing	MAE Value	MdSE Value
After one	RF	0.863%	0.8841%	0.981	0.599	0.0742	0.0531
month	MLP	0.872%	0.8468%	0.932	0.906	0.00070	0.0043
After three months	RF	0.891%	0.8640%	0.682	0.706	0.00179	0.0044
	MLP	0.882%	0.8792%	0.282	0.684	0.01152	0.0639
After six	RF	0.931%	0.9140%	0.782	0.606	0.00179	0.0044
	MLP	0.967%	0.9522%	0.882	0.667	0.0132	0.0639

Residuals for MLPRegressor Q-Q plot Train $R^2 = 0.882$ Prediction Error for MLPRegress Test $R^2 = 0.667$ R^2 0.15 0.15 0.576 1.00 best fit identity 0.95 0.10 0.10 0.90 0.05 0.05 als 0.85 ~ 0.00 0.00 0.80 0.75 -0.05 -0.05 0.70 -0.10 -0.10 0.65 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00 0.7 1.0 1.1 0.8 0.9 0 redicted Value Theoretical quantile (a) (b)

Figure 5. Evaluation of MLP after 6 months (**a**) residuals for MLP regressor model (**b**) prediction for MLP regressor model.

7.4. Progression Statistical Analyses

We performed many statistical tests on different parameters in the dataset to obtain statistically significant changes. We analyzed the changes in illness phases for 6 months, and they were distinguished by a statistically significant difference in CSA. When comparing severe cases to mild and moderate cases, but not between mild and moderate cases, pairwise comparisons showed that the CSA change was considerably smaller in severe cases. The analysis results are shown in detail in Table 8.

We found a statistically significant difference in CSA change over the 6 months across the three US stages (p value < 0.001). Further pairwise comparisons demonstrated that CSA change was significantly lower in severe cases than in mild and moderate cases, whilst no significant difference was observed between mild and moderate cases. The median ranges were found to be 1.1, 0.7 and 0.4 in mild, moderate, and severe stages, respectively. Therefore, we suggest that US staging can serve as a predictive tool for identifying individuals with mild and moderate CTS US stages who may respond better to hydro dissection.

Mild	Moderate	Severe	× 1 71
25%	30%	45%	- [•] <i>p</i> value
1.1	0.7	0.4	
0.9–1.4	0.6-0.74	0.4–0.6	< 0.001
А	А	В	_
	Mild 25% 1.1 0.9–1.4 A	Mild Moderate 25% 30% 1.1 0.7 0.9–1.4 0.6–0.74 A A	Mild Moderate Severe 25% 30% 45% 1.1 0.7 0.4 0.9–1.4 0.6–0.74 0.4–0.6 A A B

Table 8. Changing in CSA according to initial Ultrasound stage.

We utilized the significance test to track changes in SSS, FSS and CTS during the study period. As shown in Table 9 and Figure 6, statistically significant changes in CSA, SSS and FSS were observed in CTS cases over time (Initial > 1 month > 3 months > 6 months). Our study revealed a significant decrease in symptoms, which is evident in the SSS, FSS and pain analog scale, and a diminished CSA of the MN at 1, 3 and 6 months of postinjection compared with a baseline assessment. The CSA was lowered by approximately 1.5 msec in the first month, 1.3 msec in the third month and 1 mmseq in the sixth month from the initial value.

Table 9. A significant change in CSA, SSS, FSS.

Measurement	Initial	One-Month	Three- Months	Six-Months	F	* p	Partial η ²
			C	ĊSA			
Mean	16.6	15	15.4	15.7	12.913		
SD	4.2	3.6	3.7	3.8		< 0.001	0.249
** Pairwise	А	В	С	D		_	
			S	SSS			
Mean	36.8	21.7	25	27.7	199.018		
SD	5.5	6.6	6.5	8.3		< 0.001	0.866
** Pairwise	А	В	С	D		_	
	FSS						
Mean	23.9	13.4	14.8	16.3	197.840		
SD	3.9	3.3	3.9	4.5		< 0.001	0.855
** Pairwise	А	В	С	D			

* *p* value: One-Way repeated measures ANOVA. ** Pairwise comparisons: Similar letters = Insignificant difference, Different letters = Significant difference.

Upon examining simple main effects for US staging, no significant difference was observed in the initial stages of mild, moderate, and severe. However, after 1 month, 3 months and 6 months, the FSS and SSS were significantly higher in the severe stage than in the mild and moderate stages. The mild and moderate groups showed equal improvement in FSS and SSS after 1, 3 and 6 months, with a better chance than the severe stage. The results are shown in Table 10 and Figure 7.

7.5. Discussion

CTS is considered the most prevalent mononeuropathy caused by nerve entrapment. Treatment of CTS varies according to the initial state of the patient. Treatment can range from medical treatment to surgical operations. In recent years, the injection process has shown promising results in CTS treatment. The injectate material varies (e.g., steroid, 5% dextrose water, platelet-rich plasma, and saline), as well as the injection volumes (from 1 mL to 5 mL). Some studies have suggested that a higher injectate volume of 4 mL may

produce a better reaction [50]. In this study, we explored the effect of a 5 mL injection consisting of 2 mL dexamethasone, 1 mL lidocaine and 2 mL saline. The classification and regression models utilized in the study showed the following findings. Firstly, DL outperformed ML in tracking and predicting the progression of CTS during the treatment process. The DL model demonstrated superior performance compared to a previous ML model used in our previous study [35]. DNN model predicts the prognosis of CTS and then the ML model predicts CSA after 1, 3 and 6 months with improved accuracy. In addition, a DNN model was developed to predict the prognosis of CTS. This model utilized advanced deep learning techniques to analyze the input data and make predictions about the future course of the condition. The DNN model showed promise in providing valuable insights into the prognosis of CTS. Moreover, an ML model was built to predict CSA after specific time intervals (1, 3, and 6 months). This ML model, which may have utilized traditional machine learning algorithms, demonstrated the ability to forecast the change in CSA over time. This prediction can aid in understanding the progression of CTS and monitoring the effectiveness of treatment interventions. Moreover, the study investigated the patient's status after undergoing an injection process. Remarkably, there was a highly significant reduction in symptoms as evidenced by improvements in Symptom Severity Scale (SSS) and Functional Status Scale (FSS) scores. Additionally, CSA measurements exhibited a consistent decrease over 1, 3, and 6 months post-injection, indicating an improvement in nerve compression. The CSA reduction was approximately 1.5 msec in the first month, 1.3 msec in the third month, and 1 msec in the sixth month. These findings demonstrate the efficacy of the injection process in alleviating symptoms and reducing nerve compression.

Furthermore, the study identified a statistically significant increase in CSA in CTS cases compared to control subjects, with a cut-off point of 11 mm2. This measurement of CSA proved to be a reliable test for differentiating CTS patients from control subjects. Overall, the classification and regression models employed in this study provide valuable insights into the prognosis, prediction of CSA changes, and evaluation of treatment outcomes in CTS patients. Overall, the study's implementation of DL and ML models showcased their potential in predicting the prognosis and progression of CTS. The DL model surpassed previous ML approaches, highlighting the value of deep learning techniques in analyzing CTS data and making accurate predictions. The ML model specifically focused on predicting CSA changes, providing insights into the effectiveness of treatment over time.



Figure 6. Cont.



Figure 6. (a) Changing in CSA over time in CTS case (b) Changing in SSS over time in CTS cases (c) FSS over time in CTS cases.

FSS						
Time Point	Mild	Moderate	Severe	F	р	
Initial	А	А	А	1.310	0.323	
1-month	А	А	В	21.736	< 0.001	
3-months	А	А	В	21.873	0.002	
6-months	А	А	В	12.970	0.011	
		SS	S			
Time point	Mild	Moderate	Severe	F	р	
Initial	А	А	А	5.093	0.052	
1-month	А	А	В	18.772	< 0.001	
3-months	А	А	В	16.745	< 0.001	
6-months	А	А	В	22.194	< 0.001	

Table 10. FSS and SSS: Simple main effect for group.



Figure 7. Profile plot showing US grouping and time two-way interaction in (a) FSS and (b) SSS.

7.6. Strengths and Limitations

This study presents strengths to the field of CTS diagnosis and treatment, including the following:

1-Aggregated CTS dataset: The dataset from Kaferelshikh University includes a substantial number of samples (160), which consist of CTS and non-CTS patients. 2-Differentiation of CTS patients: The proposed DL model has the strength to differentiate between CTS and non-CTS patients based on the BCTQ data and NCS. This model can effectively identify and classify individuals with CTS, which aids in early detection and appropriate treatment interventions.

3-Prediction of CSA: The proposed ML model can predict the CSA of the MN after 1, 3 and 6 months of postinjection. This predictive model can assist in monitoring the progress and effectiveness of treatment over time, which allows for adjustments as needed.

4-Assessment of patient health status: Through statistical tests such as ANOVA and the Kruskal-Wallis test, the study evaluated whether patient health status significantly changed after the hydro dissection injection process. This analysis provides valuable insights into the effectiveness of the treatment and its effect on patients' well-being.

Overall, these strengths highlight the utilization of a comprehensive dataset, the development of accurate prediction models and the evaluation of treatment outcomes using rigorous statistical tests. These approaches contribute to a better understanding of CTS and facilitate more informed decision making in its diagnosis and management.

One limitation of this study was the relatively small sample size of CTS patients. Although the dataset from Kaferelshikh University included 80 CTS patients, increasing the sample size would have provided a more comprehensive representation of the CTS population. A larger sample size would have allowed for more robust statistical analyses and potentially enhanced the generalizability of the findings. Future studies should aim to include a larger number of CTS patients for strengthening the validity and reliability of the results.

8. Comparison with Other Works

The use of DL has demonstrated potential in the classification and diagnosis of several diseases, including CTS. DL algorithms, with the help of complex neural networks, can identify patterns in medical images, signals, and data to predict the presence and severity of diseases. DL was used to analyze ultrasound images and EMG data for classifying and diagnosing CTS with high accuracy. For example, a DNN was used in [51] to diagnose CTS based on 415 MRI images with an accuracy of 0.63, but MRI images are expensive and may be unavailable. In [18], the authors used MNT-DeepSL based on a sample size of 84 [50(+), 34(-)] and obtained an accuracy of 0.9, but the number of cases used for analysis was small for the dataset, which may result in a less robust model. The size of the data was also small in [12,19]. In [11], the authors used XGB with a sample size of 1073 [254(+), 761(-)] and obtained an accuracy of 76.6. DL algorithms were used in [13] to predict functional and symptomatic improvement after carpal tunnel decompression surgery based on QuickDASH response data with a sample size of 1916 and obtained an accuracy of 0.72. Our proposed model utilized bagging with Adam optimizers in 160 [80(+), 80(-)], which achieved 0.969 and 0.972 in terms of ACC and AUC, respectively. Table 11 details the comparison with other studies in CTS classification. To ensure a fair comparison, we conducted a test of our model on another CTS dataset [52]. We aimed to select a dataset that encompassed similar features to those we relied on in our research. However, the performance of our model on this dataset was found to be lower than our previously obtained results. This outcome provides further confirmation that the superior performance of our proposed model can be attributed to the careful selection of features that have a significant impact on the classification process. We applied Adam Optimizer, and the model achieved several metrics (ACC = 0.809%, precision = 0.792%, R = 0.817%, F-measure = 0.802% and AUC = 0.811%). Figure 8 show the model accuracy and model loss of the model with other data. Our proposed model outperforms the state-of-the-art models for several reasons: (1) Previous studies have focused mainly on the distinction between patients with CTS and those without the condition that cannot be aligned with medical considerations. By contrast, our study gathered data from patients with CTS and other conditions that share overlapping symptoms, including cervical radiculopathy, de Quervain tendinopathy and peripheral neuropathy. (2) Our model for CTS diagnosis

incorporates historical data that significantly affects the accuracy of disease identification. (3) A regression model is provided to predict CSA after 1, 3 and 6 months for determining progression during treatment.

Reference	Models	Dataset	Results	Туре	Data Availability	
[45]	Deep CTS	415 patients	Accuracy: 0.63	MRI	Private	
[18]	MNT-DeepSL	84 patients	Accuracy: 0.9	US	Private	
[19]	SVM	65 patients	Accuracy: 0.901	US	Private	
[11]	XGB	1037 patients	Accuracy: 76.6	EDx	Public	
[12]	SVM	38 patients	Accuracy: 0.9513	EDx	Private	
[13]	QuickDASH	1916 patients	Accuracy: 0.72	BCTQ	Public	
	Proposed	160 patients	Accuracy: 0.969	US, EDx BCTO	Private	

Table 11. Compared to previous research in terms of CTS classification.



Figure 8. Evaluation of DL model with other data set (a) model accuracy (b) model loss data.

Table 12 shows comparisons with other studies in predicting CTS. Very few works focus on the prediction of CSA after 1, 3 and 6 months based on ML models. All these studies have developed their models based on data collected after the patient has undergone surgery. For example, gradient boosting was utilized in [13] to forecast the likelihood of post surgery improvement by aggregating data from 2119 patients. The study reported an AUC of 0.7820 for the ability of their model to predict patient progress. In [15], the authors used Mask R-CNN with a sample size of 103. This model predicts the CSA automatically calculated from the MN section. The proposed model achieved promising results in terms of different prediction metrics (DSC: 0.86, precision: 0.86). In [17], the authors used Efficient Net with a sample size of 100 and achieved an accuracy of 0.93. The proposed model (MLP) achieved promising results in terms of different metrics to predict after 1 month (ACC = 0.8468, MdSE = 0.0043), after 3 months (ACC = 0.8792, MdSE = 0.0639) and after 6 months (ACC = 0.9522, MdSE = 0.0639). Accordingly, our studies demonstrate the potential of ML models to accurately predict CSA changes in CTS patients after various treatment durations. This information can be valuable for clinicians in monitoring treatment response and adjusting treatment plans as needed.

Reference	Models	Dataset	Evaluation Measures
[13]	Gradient boosting	2119 patients	AUC: 0.7820
[15]	Mask R-CNN	103 patients	Precision: 0.86 DSC 0.86
[17]	Efficient Net	100 patients	Accuracy: 0.93
	Proposed	160 patients	Accuracy: 0.9522

Table 12. Compared to previous research in terms of predicting diagnosis.

9. Model Explanation

In light of the promising results of our developed model, there remain concerns regarding its reliability when viewed through the lens of a medical expert. For this purpose, we have selected the ensemble classifier that demonstrated the highest accuracy. To interpret this model, we employ the SHAP explainer, which offers both global and local explanations.

Figure 9 showcases the SHAP values, which indicate the feature importance according to the model. The y-axis represents the features, while the x-axis represents the impact of each feature. The most significant features are located at the top, with blue and red bars denoting their contribution to the positive and negative classes, respectively. Notably, FSSI, SQSI and TINNEL as the most influential features, exerting an equal impact on both classes. To gain deeper insights into individual instances, Figure 10 portrays the average impact of each feature on each instance, and the force plot provides local explanations. Feature names are displayed on the x-axis, while the length of each bar represents the feature's importance in terms of the instance values. The force plot allows us to trace the cumulative contributions lowering it. Notably, in Figure 10 CSA, TINNEL and SAS3 exert a negative influence on the prediction, push the prediction towards the negative class according to the values of the instance. These findings offer valuable insights into the decision-making process of the model, validating the influence of specific features on the predictions and aligning with existing medical research.



Figure 9. Summary plot of the proposed model.



Figure 10. Force plot of the proposed model.

10. Conclusions and Future Work

In this study, we conducted an investigation into the diagnosis, prediction, and treatment monitoring of Carpal Tunnel Syndrome (CTS) using Deep Learning (DL) and Machine Learning (ML) models. Our study encompassed a cohort of 160 patients, comprising individuals with varying degrees of CTS severity as well as non-CTS patients who exhibited similar disease symptoms. Our findings demonstrated that DL models exhibited a remarkable level of accuracy in diagnosing CTS, with the best-performing model achieving an accuracy of 96.9%, a precision of 98.2%, and a recall of 96.3%. This indicates the efficacy of DL models in accurately identifying CTS cases. Furthermore, the ML models demonstrated excellent predictive capabilities for measuring Cross-Sectional Area (CSA) changes after 1, 3, and 6 months. The top-performing ML model achieved an accuracy (ACC) of 95.22%, an R2 score of 0.667, a Mean Absolute Error (MAE) of 0.0132, and a Median Squared Error (MdSE) of 0.0639. These results highlight the ML models' ability to accurately forecast CSA alterations over time. Statistical tests conducted in our study revealed a highly significant reduction in symptoms and CSA after 1, 3, and 6 months of post-injection treatment. This provides strong evidence for the effectiveness of the employed treatment approach. Based on our research outcomes, we recommend that future studies should focus on increasing the sample size of CTS patients to enhance the generalizability of the findings. Additionally, we suggest implementing our developed DL and ML models within the Department of Neuroscience at Kafrelsheikh University, as they have demonstrated promising results in CTS diagnosis, prediction, and treatment monitoring.

Author Contributions: Conceptualization, N.E.-R., R.R.M., A.A., A.E., M.E. and H.E.-B.; methodology, M.E. and N.E.-R., A.E.; software, M.E. and N.E.-R.; validation, A.A., H.E.-B., R.R.M. and A.E.; formal analysis, M.E. and N.E.-R.; investigation, M.E. and K.A.; resources, M.E. and S.E., A.E.; data curation, M.E. and S.E.; writing—original draft preparation, M.E. and K.A.; writing—review and editing, M.E., N.E.-R. and K.A.; visualization, M.E. and R.R.M.; supervision, R.R.M., A.A., A.E., K.A. and H.E.-B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study did not require ethical approval. The dataset was obtained retrospectively from the Neurology department at Kaferelshikh University Hospitals in Egypt (Approval code: MKSU 50-12-27).

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available upon request.

Acknowledgments: This study is supported via funding from Prince sattam bin Abdulaziz University project number (PSAU/2023/R/1445).

Conflicts of Interest: The authors declare no conflict of interest

Abbreviations

The following	ng abbreviations are used in this manuscript:		
CTS	Carpal Tunnel Syndrome		
US	Ultrasound		
NCS	Nerve Conduction Studies		
CSA	Cross-Sectional Area		
EDx	electrodiagnostic		
SVM	Support Vector Machines		
DT	Decision Trees		
XGB	Extreme gradient boosting		
RF	Random Forest		
LR	Logistic Regression		
XGB	eXtreme Gradient Boosting		
KNN	k-Nearest Neighbors		
NB	Naive–Bayes		
NN	Neural Network		
SGB	Stochastic Gradient Boosting		
MRI	Magnetic resonance image		
MLP	Multilayer Perception		
ROI	region of interest		
MNT	median nerve localization		
BMI	Body mass index		
BCTQ	Boston Carpal Tunnel Syndrome Questionnaire		
FSS	functional status scale		
SSS	symptoms severity scale		
MAE	Mean Absolute Error		
MdSE	Median Squared Error		
ACC	Accuracy		
R2	score (coefficient of determination)		
DSC	Dice Similarity Coefficient		
IoU	intersection over union		
PROMs	patient-reported outcome measure		
GD	Gradient descent		
Adagrad	Adaptive Gradient Algorithm		

References

- Levine, D.W.; Simmons, B.P.; Koris, M.J.; Daltroy, L.H.; Hohl, G.G.; Fossel, A.H.; Katz, J.N. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. *J. Bone Joint Surg. Am.* 1993, 75, 1585–1592. [CrossRef] [PubMed]
- 2. Pastare, D.; Therimadasamy, A.K.; Lee, E.; Wilder-Smith, E.P. Sonography versus nerve conduction studies in patients referred with a clinical diagnosis of carpal tunnel syndrome. *J. Clin. Ultrasound* **2009**, *37*, 389–393. [CrossRef]
- Stecco, C.; Giordani, F.; Fan, C.; Biz, C.; Pirri, C.; Frigo, A.C.; Fede, C.; Macchi, V.; Masiero, S.; De Caro, R. Role of fasciae around the median nerve in pathogenesis of carpal tunnel syndrome: Microscopic and ultrasound study. J. Anat. 2020, 236, 660–667. [CrossRef] [PubMed]
- Phalen, G.S. The Carpal-Tunnel Syndrome: Seventeen years' experience in diagnosis and treatment of six hundred fifty-four hands. *JBJS* 1966, 48, 211–228. [CrossRef]
- Awad, M.; Khanna, R. Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers; Springer Nature: Berlin, Germany, 2015; 248p. [CrossRef]
- El-Bakry, H.M.; Mastorakis, N. A new fast forecasting technique using high speed neural networks. WSEAS Trans. Signal Process. 2008, 4, 573–581.
- 7. Srikanteswara, P.K.; Cheluvaiah, J.D.; Agadi, J.B.; Nagaraj, K. The relationship between nerve conduction study and clinical grading of carpal tunnel syndrome. *J. Clin. Diagnostic Res.* **2016**, *10*, 13–18. [CrossRef]
- El-Bakry, H.M.; Mastorakis, N. Fast human motion tracking by using high speed neural networks. In Proceedings of the 9th WSEAS International Conference on Signal, speech and image processing, and 9th WSEAS international conference on Multimedia, Internet & Video Technologies, Budapest, Hungary, 3–5 September 2009; pp. 221–240.
- 9. El-Bakry, H.M.; Zhao, Q. Speeding-up Normalized Neural Networks for Face/Object Detection. MG&V 2005, 14, 29–59.

- Hoogendam, L.; Bakx, J.A.C.; Souer, J.S.; Slijper, H.P.; Andrinopoulou, E.R.; Selles, R.W. Predicting Clinically Relevant Patient-Reported Symptom Improvement After Carpal Tunnel Release: A Machine Learning Approach. *Neurosurgery* 2022, 90, 106–113. [CrossRef]
- 11. Park, D.; Kim, B.H.; Lee, S.E.; Kim, D.Y.; Kim, M.; Kwon, H.D.; Kim, M.C.; Kim, A.R.; Kim, H.S.; Lee, J.W. Machine learning-based approach for disease severity classification of carpal tunnel syndrome. *Sci. Rep.* **2021**, *11*, 17464. [CrossRef]
- 12. Tsamis, K.I.; Kontogiannis, P.; Gourgiotis, I.; Ntabos, S.; Sarmas, I.; Manis, G. Automatic electrodiagnosis of carpal tunnel syndrome using machine learning. *Bioengineering* **2021**, *8*, 181. [CrossRef]
- Harrison, C.J.; Geoghegan, L.; Sidey-Gibbons, C.J.; Stirling, P.H.C.; McEachan, J.E.; Rodrigues, J.N. Developing Machine Learning Algorithms to Support Patient-centered, Value-based Carpal Tunnel Decompression Surgery. *Plast. Reconstr. Surg. Glob. Open* 2022, 10, e4279. [CrossRef] [PubMed]
- 14. Ciobanu, D.M.; Stan, A.D.; Nicu, C.; Leucut, D. Clinical Utility of Boston-CTS and Six-Item CTS Questionnaires in Carpal Tunnel Syndrome Associated with Diabetic Polyneuropathy. *Diagnostics* **2023**, *13*, 4.
- Smerilli, G.; Cipolletta, E.; Sartini, G.; Moscioni, E.; Di Cosmo, M.; Fiorentino, M.C.; Moscia, S.; Frontoni, E.; Grassi, W.; Filippucci, E. Development of a convolutional neural network for the identification and the measurement of the median nerve on ultrasound images acquired at carpal tunnel level. *Arthritis Res. Ther.* 2022, 24, 38. [CrossRef] [PubMed]
- Di Cosmo, M.; Chiara Fiorentino, M.; Villani, F.P.; Sartini, G.; Smerilli, G.; Filippucci, E.; Frontoni, E.; Moccia, S. Learning-Based Median Nerve Segmentation from Ultrasound Images for Carpal Tunnel Syndrome Evaluation. In Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Virtual, 1–5 November 2021; pp. 3025–3028. [CrossRef]
- Shinohara, I.; Inui, A.; Mifune, Y.; Nishimoto, H.; Yamaura, K.; Mukohara, S.; Yoshikawa, T.; Kato, T.; Furukawa, T.; Hoshino, Y.; et al. Using deep learning for ultrasound images to diagnose carpal tunnel syndrome with high accuracy. *Ultrasound Med. Biol.* 2022, 48, 2052–2059. [CrossRef]
- 18. Wang, Y.W.; Chang, R.F.; Horng, Y.S.; Chen, C.J. MNT-DeepSL: Median nerve tracking from carpal tunnel ultrasound images with deep similarity learning and analysis on continuous wrist motions. *Comput. Med. Imaging Graph.* **2020**, *80*, 101687. [CrossRef]
- Faeghi, F.; Ardakani, A.A.; Acharya, U.R.; Mirza-Aghazadeh-Attari, M.; Abolghasemi, J.; Ejtehadifar, S.; Mohammadi, A. Accurate automated diagnosis of carpal tunnel syndrome using radiomics features with ultrasound images: A comparison with radiologists' assessment. *Eur. J. Radiol.* 2021, 136, 109518. [CrossRef] [PubMed]
- 20. Hafiane, A.; Vieyres, P.; Delbos, A. Deep learning with spatiotemporal consistency for nerve segmentation in ultrasound images. *arXiv* 2017, arXiv:1706.05870.
- 21. Lee, K.J. Architecture of neural processing unit for deep neural networks. Adv. Comput. 2020, 122, 217–245. [CrossRef]
- 22. Hassan, E.; Talaa, N.E.F.M. Review : Mask R—CNN Models. Nile J. Commun. Comput. Sci. 2022, 3, 1–10. [CrossRef]
- 23. Shanmugavadivu, P.; Shanthi Rani, M.; Chitra, P.; Lakshmanan, S.; Nagaraja, P.; Vignesh, U. Bio-Optimization of Deep Learning Network Architectures. *Secur. Commun. Networks* 2022, 2022, 3718340. [CrossRef]
- Hassan, E.; Elmougy, S.; Ibraheem, M.R.; Hossain, M.S.; AlMutib, K.; Ghoneim, A.; AlQahtani, S.A.; Talaat, F.M. Enhanced Deep Learning Model for Classification of Retinal Optical Coherence Tomography Images. *Sensors* 2023, 23, 5393. [CrossRef] [PubMed]
- Ostertagová, E.; Ostertag, O. Methodology and Application of Oneway ANOVA. Am. J. Mech. Eng. 2013, 1, 256–261. [CrossRef]
 Siegel, A.F. Chapter 15—ANOVA: Testing for Differences Among Many Samples and Much More, 7th ed.; Academic Press: Cambridge, MA, USA, 2016; pp. 469–492. [CrossRef]
- 27. Sedgwick, P. Understanding P values. BMJ 2014, 349, g4550. [CrossRef]
- 28. Dahiru, T. P-value, A true test of statistical significance? A cautionary note. Ann. Ibadan Postgrad. Med. 2008, 6, 21–26. [CrossRef]
- 29. Gurevich, Y.; Vovk, V. Test statistics and p-values. Proc. Mach. Learn. Res. 2019, 105, 89–104.
- 30. McKight, P.; Najab, J. Kruskal-Wallis Test. Corsini Encycl. Psychol. 2010, 1, 1–10. [CrossRef]
- 31. Padua, L.; Coraci, D.; Erra, C.; Pazzaglia, C.; Paolasso, I.; Loreti, C.; Caliandro, P.; Hobson-Webb, L.D. Carpal tunnel syndrome: Clinical features, diagnosis, and management. *Lancet Neurol.* **2016**, *15*, 1273–1284. [CrossRef] [PubMed]
- 32. Alfonso, C.; Jann, S.; Massa, R.; Torreggiani, A. Diagnosis, treatment and follow-up of the carpal tunnel syndrome: A review. *Neurol. Sci.* 2010, *31*, 243–252. [CrossRef]
- 33. Chandy, B.R. Ultrasonograpghy—A Diagnostic Aid for Carpal Tunnel Syndrome. Ph.D. Thesis, Christian Medical College, Vellore, India, 2009.
- Fischer, J.; Thompson, N.W.; Harrison, J.W.K. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. In *Classic Papers in Orthopaedics*; Springer: London, UK, 2014; pp. 349–351. ISBN 9781447154518.
- 35. Elseddik, M.; Mostafa, R.R.; Elashry, A.; El-Rashidy, N.; El-Sappagh, S.; Elgamal, S.; Aboelfetouh, A.; El-Bakry, H. Predicting CTS Diagnosis and Prognosis Based on Machine Learning Techniques. *Diagnostics* **2023**, *13*, 492. [CrossRef]
- Velázquez, F.; Berná, J.D.; Abellán, J.L.; Serrano, L.; Escribano, A.; Canteras, M. Reproducibility of sonographic measurements of carotid intima-media thickness. *Acta Radiol.* 2008, 49, 1162–1166. [CrossRef] [PubMed]
- 37. El Miedany, Y.M.; Aty, S.A.; Ashour, S. Ultrasonography versus nerve conduction study in patients with carpal tunnel syndrome: Substantive or complementary tests? *Rheumatology* **2004**, *43*, 887–895. [CrossRef]

- Alsaeid, M.A. Dexamethasone versus Hyaluronidase as an Adjuvant to Local Anesthetics in the Ultrasound-guided Hydrodissection of the Median Nerve for the Treatment of Carpal Tunnel Syndrome Patients. *Anesth. essays Res.* 2019, 13, 417–422. [CrossRef] [PubMed]
- 39. Azur, M.J.; Stuart, E.A.; Frangakis, C.; Leaf, P.J.; Washington, D.C. Multiple imputation by chained equations: What is it and how does it work ? *Int. J. Methods Psychiatr. Res.* 2011, 20, 40–49. [CrossRef]
- 40. Zhang, Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. *Ann. Transl. Med.* **2016**, *4*, 30. [CrossRef] [PubMed]
- 41. Buuren, S.; Groothuis-Oudshoorn, C. MICE: Multivariate Imputation by Chained Equations in R. J. Stat. Softw. 2011, 45, 1–67. [CrossRef]
- 42. Sharma, V. A Study on Data Scaling Methods for Machine Learning. Int. J. Glob. Acad. Sci. Res. 2022, 1, 23–33. [CrossRef]
- 43. Gamel, S.A.; Hassan, E.; El-Rashidy, N.; Talaat, F.M. Exploring the effects of pandemics on transportation through correlations and deep learning techniques. *Multimed. Tools Appl.* **2023**, 1–22. [CrossRef]
- 44. El-Bakry, H.M. An efficient algorithm for pattern detection using combined classifiers and data fusion. *Inf. Fusion* **2010**, *11*, 133–148. [CrossRef]
- 45. El-Bakry, H.; Mastorakis, N. New Fast Normalized Neural Networks for Pattern Detection. *Image Vis. Comput. J.* 2007, 25, 1767–1784. [CrossRef]
- El-Bakry, H.; Abo-elsoud, M.; Kamel, M. Fast Modular Neural Networks for Human Face Detection. In Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, Como, Italy, 24–27 July 2000; pp. 320–324.
- El-Bakry, H. Fast Iris Detection for Personal Verification Using Modular Neural Networks. In Proceedings of the 7th Fuzzy Days International Conference, Dortmund, Germany, 1–3 October 2001; pp. 269–283.
- El-Bakry, H. Comments on using MLP and FFT for fast object/face detection. In Proceedings of the International Joint Conference on Neural Networks, Portland, OR, USA, 20–24 July 2003. [CrossRef]
- 49. El-Bakry, H.M. Fast virus detection by using high speed time delay neural networks. J. Comput. Virol. 2009, 6, 115–122. [CrossRef]
- Lin, M.-T.; Liao, C.-L.; Hsiao, M.-Y.; Hsueh, H.-W.; Chao, C.-C.; Wu, C.-H. Volume Matters in Ultrasound-Guided Perineural Dextrose Injection for Carpal Tunnel Syndrome: A Randomized, Double-Blinded, Three-Arm Trial. *Front. Pharmacol.* 2020, 11, 62583. [CrossRef] [PubMed]
- 51. Zhou, H.; Bai, Q.; Hu, X.; Alhaskawi, A.; Dong, Y.; Wang, Z.; Qi, B.; Fang, J.; Kota, V.G.; Abdulla, M.H.A.H.; et al. Deep CTS: A Deep Neural Network for Identification MRI of Carpal Tunnel Syndrome. J. Digit. Imaging 2022, 35, 1433–1444. [CrossRef] [PubMed]
- Czerniak, D. Carpal-Tunnel-Syndrome. Available online: https://github.com/polemizatorr/Classification-of-carpal-tunnelsyndrome (accessed on 14 July 2023).

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.