بسم الله الرحمن الرحيم

Imam Reza International University

Master's Thesis

Diagnosing Parkinson's disease using handwriting lines with the help of deep convolutional neural networks

Student Name: Yasser Ali Khaled Al-Nuaimi

Professor Name: Dr. Saeed Sherbaf

Year 2010

Abstract

Parkinson's disease cannot be diagnosed and identified until the onset of motor symptoms and signs. Therefore, early detection of this disease is very crucial to prevent adverse events.

This study deals with the diagnosis of Parkinson's disease based on handwriting patterns. The proposed algorithm has several steps; including image processing, feature extraction, and classification. One of the most important steps in this process is the implementation of the classifier. More power in the learning process of the classifier leads to enhanced detection accuracy. For this purpose, we used the CNN classifier to classify and diagnose the disease and achieved a detection accuracy of 80.4%.

Keywords: Parkinson, Handwriting, Convolution neural network.

Contents

List figures

Chapter One: Introduction and Overview of the Research

1-1 Introduction

Today, most organizations store a huge amount of data and information in their databases. These organizations need to understand such data and extract useful and existing knowledge. Data mining, as an interdisciplinary science, can reveal valuable and hidden information in a huge amount of data. Data mining integrates various fields to do this, such as database, statistics, machine learning, and other related fields [14].

In fact, the process of recognizing models and patterns in data is called knowledge discovery. The characteristics of these models and patterns include usefulness, validity, novelty, and complete understanding. A stage of the knowledge discovery process during which patterns are found in data is called data mining. These patterns are obtained based on specific data mining methods and with computational efficiency [14].

In other words, data mining extracts knowledge from big data and performs the necessary analysis, investigation, and prediction. Essentially, this process goes beyond data collection and management. By providing the ability to recognize trends in data and also automatically extract patterns, data mining is used in various industries and sectors such as insurance, banking, medicine, and retail with the goals of strengthening research, increasing sales, and reducing costs. In the past few decades, the use of data mining in the medical field has been welcomed. The medical environment usually has a lot of information and requires knowledge discovery. Data mining techniques are used to emerge this science. The reason for the use of data mining methods by medical researchers in advanced countries is the rapid growth of medical databases. These researchers use these techniques to discover science. Today, hidden patterns in data are discovered using data mining techniques. Subsequently, such patterns are appropriate for the databases. For example, by

building relevant patterns in the medical field, a decision support system can be created to help doctors draw conclusions. One of the challenging areas in the past decade has been how to apply data mining and machine learning methods to medical and health data. The characteristics of medical data mining are specific to this field. For this reason, data mining in this field is different from data mining in other problems and applications. For example, researchers using data mining in other fields may encounter fewer problems and limitations. Problems such as heterogeneous data, private or confidential data, distributed and large data, as well as ethical, legal, and security aspects should also be considered [13].

In the field of handling big data, data mining expands and is integrated with methods and fields such as machine learning, pattern recognition, statistics, artificial intelligence, and database. Therefore, many techniques have been inspired by these sciences [11].

1-2 Problem Statement

Research conducted in the field of medicine has proven that the handwriting of Parkinson's patients is different from that of normal individuals. Researchers want to obtain parameters to well represent these changes. In the research conducted, distinct methods and solutions were investigated and presented for this purpose. The aim of the solutions is to provide a suitable proposal for processing handwritings in order to diagnose and identify Parkinson's disease. In recognizing and examining handwriting patterns, the neural network classifier plays an important and prominent role in classification. A neural network is an approximate model of the brain and, like connected neurons in the brain, has a parallel structure and function. A neural network has the ability to learn and adapt. It also has the power to adjust parameters. Support vector machine (SVM) is one of the best-known and most widely used methods in the field of classification. Input vectors and parameters are mapped to a

multidimensional space. Then, a hyper-plane is constructed and separates the inputs from each other with the greatest distance. The goal of this method is to determine decision boundaries in such a way that the minimum distance to each class is maximum. The boundary determination method is based on support vectors. Data close to the separating hyper-planes are called support vectors [1,2,3].

1-3 Importance of Research

Parkinson's disease cannot be diagnosed and identified until motor symptoms and signs appear. Therefore, early diagnosis of this disease is very crucial in order to prevent adverse events. For various reasons, such as the advanced age of patients and motor problems, the identification of this disease should not require the presence of the individual patient in the clinic. Also, the information should be easily transferable [4].

1-4 Applications of medical data mining

Much of the knowledge accumulated in medical databases is the result of modern medical processes. The most important applications of data mining in the fields of medicine and healthcare are introduced below. Data mining has many and diverse applications in these fields [13].

1-4-1 Prediction and diagnosis of diseases

Today, a large amount of information about patients and their medical conditions is collected in clinical databases. Useful knowledge can be discovered in such data sets and used in scientific decision-making for analyzing and diagnosing diseases. Diagnosis of disease and appropriate treatment method in medical science is of particular importance for patients. If appropriate treatment is not applied to the patient, it has destructive effects and causes waste of time and money. Also, incorrect selection in some cases leads to the death of the patient. Some doctors do not have sufficient skills in their professional field. In some cases, there are few experts. As

a result, the existence of a system for diagnosing and selecting the appropriate type of treatment for patients is useful. Such systems also lead to cost reduction in addition to increasing accuracy and efficiency. One of the most important causes of death in most countries is chronic diseases such as diabetes, cardiovascular diseases, and obesity. By searching and exploring the data of similar patients in the past and also comparing and matching common symptoms, it is possible to diagnose and predict diseases [13].

1-4-2 Effectiveness of treatments
Data mining provides an accurate analysis of the path of successful operations by comparing the factors, symptoms, and treatment methods of a disease. The results of such studies can be used to evaluate the effectiveness of medical treatments. For example, the best treatment method with the lowest cost and complications can be determined by comparing the results of classifying patients treated with different drug groups for a specific disease. It is also possible to identify successful treatments for a specific disease based on data mining. Examples of data mining applications in this field include:

- Identifying the relationship between different side effects of a treatment method
- The relationship between the patient's treatment and response
- Determining the most suitable drug combinations for treating a subset of patients with a specific disease, especially different reactions to prescribed drugs during treatment
- Determining preventive steps to reduce pain

1-4-3 Health and medical services management
Data mining in the field of health and medical services management is effective in activities such as identifying more widely and better chronic diseases and high-risk

patients, planning appropriate activities, and reducing requests and referrals to medical centers. For example, based on demographic characteristics, the medical status of patients can be classified. Identifying groups in greater need of medical resources also facilitates the development of programs for training and management of their medical condition. Searching and exploring data in a medical center can improve services. It can also help to improve the patient's condition faster, reduce the length of hospitalization, prevent clinical complications, and provide useful information needed by doctors, as well as predict items such as the number of manpower required for various operations, the number of clients, the amount of working hours, the number of surgeries, drugs, and equipment needed for the future.

1-5 Challenges of data mining in medicine

The implementation and use of data mining in the medical field creates problems and difficulties that distinguish it from data mining in other applications. Familiarity and awareness of these problems will certainly improve medical data mining and make it more effective in supporting decision-making and preventing errors. Below are some of the challenging issues in different fields of medical data mining.

1-5-1 Heterogeneous medical data

The specific characteristics of medical data are rawness, volume, non-uniformity, and distribution. The huge volume of data is a problem that has involved the medical industry. For example, imaging techniques such as MRI, SPECT, PET, and a group of ECG and EEG signals produce up to several gigabytes of data daily. The existence of different structures in medical data sets causes a kind of inconsistency and entanglement. Medical data is obtained from various sources, as described below.

- Various images and signals
- Patient examinations
- Laboratory and clinical findings

- Physician notes, observations, and interpretations
- Accounting and admission departments

All of the above can be effective in prediction, identification, and treatment and cannot be ignored. Storing large amounts of heterogeneous data requires high-capacity systems. Data analysis and exploration also requires powerful data mining tools. Despite the large number of data mining systems and tools, their ability to manage and utilize large and heterogeneous amounts of data varies. A very important and decisive point about this data is the ability to use and extract knowledge from it. If the information is stored in an incomprehensible format, it is less useful. In these situations, methods such as random sampling, selection of meaningful features and attributes related to the data mining objective are used to convert the data into an understandable and uniform format and to reduce the data dimensions. When faced with big data, the use of visualization methods can play an important role. By presenting the results in the form of a snapshot, this technique displays a lot of information [13].

1-5-2 Confidentiality and security of human data

One of the important principles of medicine is the confidentiality of doctors and medical centers and the trust of the patient. Therefore, researchers should pay special attention to the security of medical data. Compliance with the principle of confidentiality of information makes it impossible to directly access all data. In order to support the principle of confidentiality of medical data, certain regulations are developed by some governments. For example, it is recommended that data related to patient identification such as name, address, telephone number be hidden. According to one of these recommendations, patient identification data should not be accessible in any way. For example, only information related to determining the disease should be accessible. In some cases, this is not acceptable. The reason is that

in most cases, more information is needed from a specific patient. Also, communication with the patient is needed to verify the accuracy of data mining methods. For example, after determining the disease and treatment using data mining methods, communication with the individual is required. One of the important issues in the use and especially the transfer of data is security. Before hiding patient identification information, only authorized persons are allowed to access this data. Medical data can be prepared using one of the following four methods.

- Data without identifier: At the time of data collection, the identification information of the patients has been removed.

- Anonymized data: The identification information of the patients is included with the original data. Such information is deleted over time and irreversibly. Subsequently, the original data cannot be retrieved.

- Data with encoded identifiers: The identification information of the patients is encoded with the original data. In order to identify these patients, this information is decoded under the conditions set by the encoder.

- Identified data: With the written consent of the patients, their identification information is included in the data set. For research and data mining purposes, identifiable data is the most suitable. If necessary, under certain conditions, it is possible to access the original data. The loss of patient trust can be the only challenge with this method. As a result, to prevent unauthorized access to data, data mining experts should be aware of the existing standards. They should also be familiar with how to prepare the data to keep the information secret.
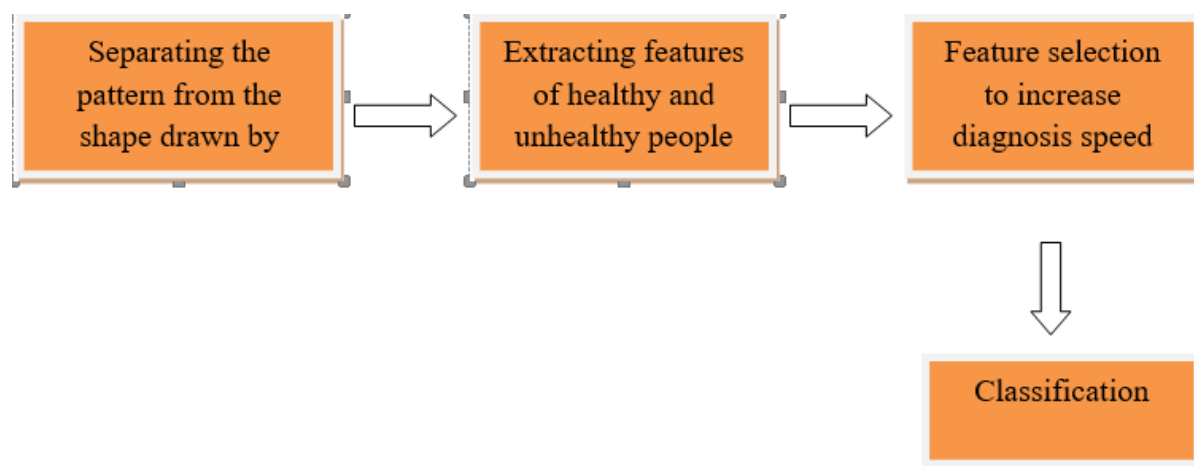
1-5-3 Need for experts

In large medical datasets, identifying patterns and determining hidden connections in this data is possible based on data mining. However, without cooperation with experts and medical professionals, these results are not useful. Some patterns

obtained from data mining are not useful and meaningful. These discovered patterns should be examined and evaluated in terms of logic, feasibility, and innovation by experienced medical professionals.

1-6 Research objectives

This study aims to provide a method and solution for diagnosing and identifying Parkinson's disease through handwriting and hand tremor patterns. To this end, this goal is achieved in the following steps.

Separating the pattern from the shape drawn by	⟹	Extracting features of healthy and unhealthy people	⟹	Feature selection to increase diagnosis speed

⟱

Classification

This method has several limitations, for example, the patient has not yet experienced a movement disorder in the early stages of the disease. The handwriting and pattern of these individuals are completely similar to healthy individuals. In some cases, the acquired pattern is similar to the data of Parkinson's patients. Therefore, the accuracy of recognizing the pattern of individuals is definitely affected.

1-7 Organization of the thesis

In this chapter, the overview, basic concepts, and also the innovation of the research in terms of studying the nervous system and handwriting production operations were described. Chapter 2 deals with the concepts and methods proposed in the diagnosis

of Parkinson's. Chapter 3 reviews some studies conducted in the field of diagnosing Parkinson's based on handwriting. Chapter 4 presents the method proposed in the leading research. Chapter 5 presents the results of the proposed method. Finally, suggestions for future work in this field will be recommended.

Chapter Two: Research Literature

2-1 Definitions of data mining

Advanced information management technology includes various types of information. The main challenge is data mining and discovering knowledge from large amounts of data. At the same time, traditional statistical methods and management tools are not enough to analyze this data. In fact, the goal of data mining operations is to extract only useful information. The exponential growth of this data worldwide multiplies the importance of this operation [15].

A brief description of the common definitions of knowledge discovery and data mining is provided below.

- In large and complex data sets, descriptive computer data is analyzed.
- In big data sets, a secondary analysis is performed.
- Structure exploration in information emphasizes the similarity of algorithmic research with research carried out with the help of an information management system.
- The process of identifying valid, novel, useful, and ultimately understandable patterns in data is knowledge discovery.
- Applying a version of machine learning to big data includes a wider range of tasks and unsupervised methods in addition to supervised learning.
- Data mining is a statistical technique in significant quantity and rate.
- Data mining is a rapidly growing interdisciplinary field used for information and valuable knowledge extraction from large amounts of data and combines various fields such as databases, statistics, machine learning and other related fields.
- Data mining is the discovery of understandable patterns and rules from large amounts of data. This process is also called knowledge discovery. The preference of the presenters is to use the term knowledge creation.

- Extracting knowledge from large amounts of data is called data mining, which finds interesting patterns in large amounts of information [14].

2-2 Data mining techniques

The main data mining techniques are divided into two classes: descriptive and predictive. The task of the descriptive class is to identify and generalize the properties of the data. The goal of descriptive methods is to discover patterns that can be interpreted by humans. Predicting the future behavior of data is one of the tasks of the predictive class. Estimation is the use of several variables in the database in order to estimate the future values of the parameters of interest [14]. Data mining methods are represented in Figure (2-1).
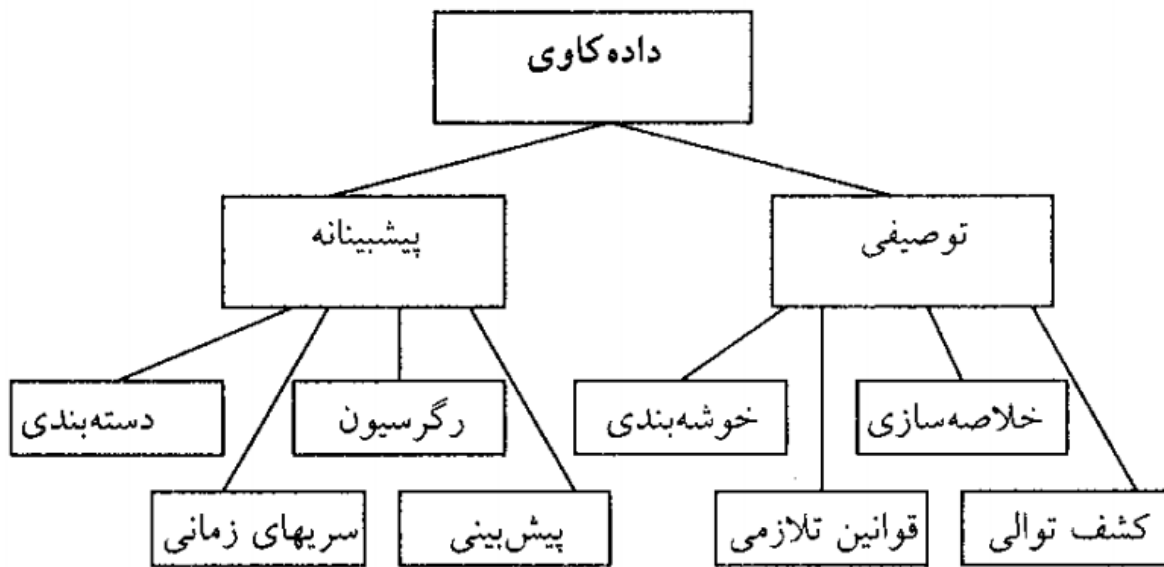

Figure 2- 1- Data mining methods [10]

2-3 Machine learning methods

In general, the analysis in this method is realized in two supervised and unsupervised ways. The main and important goal of this method is to predict t for input x. Therefore, we will discuss two different situations.

- Regression where t is a continuous variable

19

- Classification where t belongs to a continuous set

Most research falls into the second class and tries to find the best classification machine.

2-4 Support vector machine

Support vector machines or SVMs are used to detect and identify patterns and select the best solution. SVM-based classification is possible using a linear combination of the kernel function of the support vector data. Linear SVM is the easiest case of a hyper-plane to separate positive and negative samples with the greatest distance.
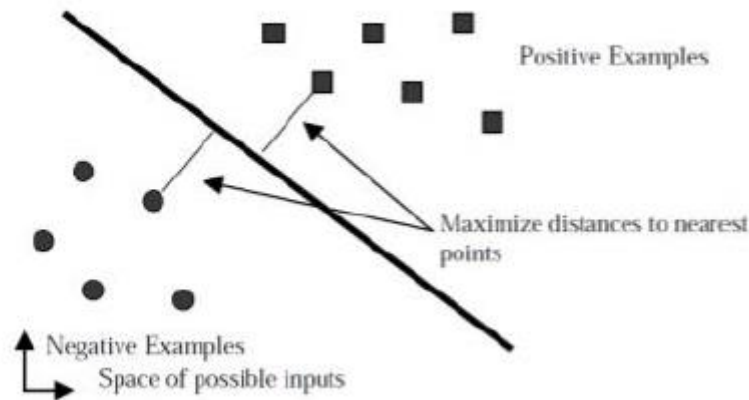

Figure 2- 2- SVM as a hyper-plane for linear separation of samples in data space

The classification performance of SVM is based on a binary separator. A combination of multi-class support vector machines can be used for pattern recognition and identification. There are generally two approaches to this goal. The first is the "one-versus-all" strategy to classify each pair of classes and the remaining classes. The second is the "one-versus-one" strategy to classify each pair [10]

2-5 Naive Bayes

In the naive Bayes structure and model, no type of relationship is allowed, and all features are assumed to be independent of each other. This structure is not used in

large-scale data classification. Recently, many efforts have been made to improve and enhance classification in areas such as selecting a set of features. Naive Bayes classification is good for most unrealistic data sets, assuming independence between features. The independence assumption ignores any relationship. Extending the structure and model of naive Bayes to clearly represent the dependence of variables is a way to overcome the limitations of naive Bayes [11].

2-6 Partitioning data into training and testing parts

In 2-fold partitioning, half of the data is assumed to be training data and the other half as testing data in a completely random manner. This process is executed twice in a row. Similarly, in 5-fold partitioning, all the data is assumed to be completely random. In the random mode, 80% of the data is considered for training and 20% for testing. In each run, the training data and testing data are swapped. Also, in 10-fold, 90% of the data is assumed for training and 10% for testing.

2-7 Convolution neural networks

Convolution neural networks (CNN) are one of the most important deep learning methods in which multiple layers are trained in a powerful way. This method is very efficient and is one of the most common methods in various computer vision applications. The architecture of a convolution neural network is shown in the figure below. In general, a CNN consists of three main layers, namely the convolution layer, the pooling layer, and the fully-connected layer. Different layers perform different tasks. In each convolution neural network, there are two stages for training, namely feed forward and backpropagation. In the first stage, the input image is fed to the network and a dot product between the input and the parameters of each neuron and finally the convolution operation is performed in each layer. Then the output of the network is calculated. In order to adjust the network parameters (network training), the output result is used to calculate the network error rate. The network

output is compared with the correct answer using an error function and the error rate is calculated. Then, based on the calculated error rate, backpropagation process begins. In this stage, the gradient of each parameter is calculated according to the chain rule and all parameters are changed according to their effect on the error generated in the network. After updating the parameters, the next stage, namely feed-forward, begins. After several iterations, the network training stages end.
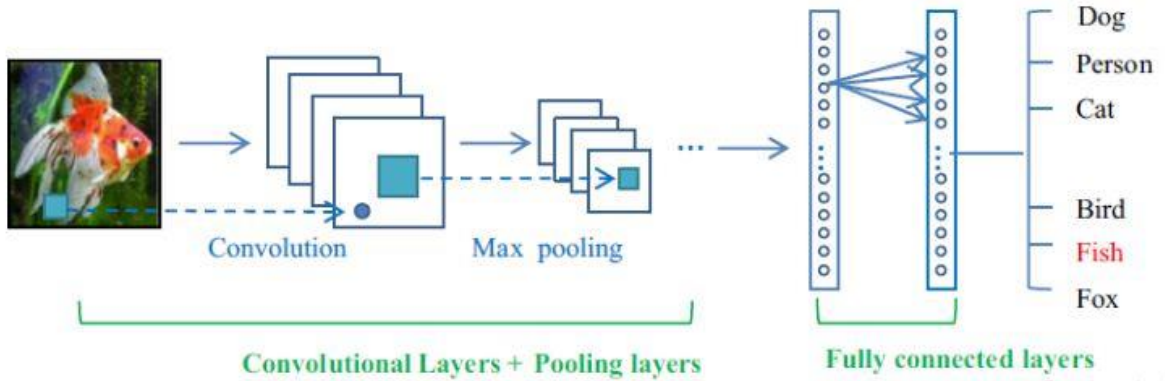

Figure 2- 3- Architecture of a convolution network

2-8 Measurement criteria

The following 3 parameters are used to measure the accuracy of the simulation and comparison with the reference paper.

$$\text{RECALL} = \frac{tp}{tp+fn}$$

$$\text{Precision} = \frac{tp}{tp+fp}$$

$$\text{F-score} = \frac{2tp}{2tp+fp+fn}$$

T_p: Indicates the people who are correctly identified as positive.

F_n: Indicates the people who are falsely identified as negative.

F_p: Indicates the people who are falsely identified as positive.

The following table represents the accuracy of the reference paper with respect to the aforementioned parameters for 3 different classifiers.

		NB	OPF	SVM
	Accuracy (%)	**78.9±3.5**	**77.1±3.8**	**75.8±3.3**
Control	F-score	0.38±0.04	0.30±0.10	0.23±0.11
	Precision	0.24±0.03	0.28±0.09	0.25±0.12
	Recall	0.91±0.09	0.32±0.12	0.24±0.14
Patient	F-score	0.49±0.10	0.82±0.03	0.83±0.03
	Precision	0.94±0.05	0.83±0.03	0.82±0.02
	Recall	0.34±0.10	0.80±0.05	0.83±0.06

2-9 Decision tree

One of the powerful classification techniques is decision tree, which is gaining popularity with the increasing growth of data mining. In decision tree-based classification, the samples are arranged in a tree from the root node to the leaf nodes [14]. The classification of the samples is done from the root downwards until it reaches the leaf nodes at the end of the path. In this path, a feature is created for each internal or non-leaf node. This feature poses a question related to the input sample. The number of branches in the tree is equal to the number of possible answers to the question at each internal node. The value of each of these branches is determined by the answer to the question. In this tree, the leaves are identified by a set of answers or a class. Since the tree represents a decision process to identify the class of the input sample, this method is known as a decision tree. When a unique answer must be presented in the form of a class or class name or in problems where the training samples are specified as pairs (value-feature) and the output of the objective function is discrete values, a decision tree can be used. For example, each example is specified as yes or no. Some examples of decision trees include C5, C4.5, ID3, and CART.

Decision trees are usually used to estimate discrete functions. The characteristics of decision trees include resistance to noise in the input data and efficiency for working with big data. For ease of understanding, the decision tree can be represented by if-then rules. It is also possible to combine the hypotheses in the decision tree in "OR" and "AND" manner. This technique can be used even in training samples that do not have all the features [16].

One of the supervised classification methods is the decision tree, which aims to reach the target variable using decision rules. The advantages of this method include simplicity, ease of understanding and elimination of decisions that do not have generalizability. Depending on the decision tree method, various criteria can be used to select nodes, such as information gain and Gini Index, DKM, and Likelihood [12].

Creating a decision tree is done in the following steps.

- Based on the gain ratio, information gain, Gini index, etc., the best feature of the training data set is selected.
- A branch is added to the tree for all selected feature values.
- The above steps are continued until the class feature is reached and placed in the leaves of the tree.
- If the tree structure is confirmed as defined, a downward movement is performed in all records (downward growth). Otherwise, the steps are started again from the beginning [17].

Information gain

Information gain is one of the most famous criteria for building a tree and selecting nodes. This measure is expressed based on the entropy criterion and its estimation equation is introduced below [12].

$$Informatin\ Gain(A) = Entropy(D) - EntropyA(D) \qquad (2\text{-}2)$$

$$Entropy(D) = -\sum_{I=1}^{c} Pi \times \log 2(Pi)$$

$$Entropy_A = \sum_{i=1}^{v} \frac{|D_1|}{D} \times Entropy(D_j)$$

2-10 Regression

The goal of regression is to predict the value of a variable using a function. This method estimates the output value based on the input values. For example, estimating the height of children by considering their age and weight values is one of the applications of regression [13].

2-11 Time series analysis

This method tests a feature at regular and equal time intervals to determine its value. For example, the values of a patient's specific features may be determined by hourly or daily tests. This method is also used to detect similarities between time intervals and predict future values [13].

2-12 Visualization

Visualization is a useful method for discovering patterns in a medical dataset. By displaying data in the form of graphs and images, it is possible to classify them into appropriate clusters. It also strengthens the power of data-based decision-making and the entire process is faster and more accurate with less cognitive effort. For example, consider a dataset of heart patients. Patients with high blood pressure have

an unusual disease state. This group is separated as a special subset from the main set to discover more knowledge by applying other data mining methods [13].

2-13 FKNN

The K-nearest neighbor algorithm uses a non-parametric structure for classification. The most common class among the k nearest neighbors is assigned to the data. FKNN is a KNN algorithm based on fuzzy theory. To detect special and unique classes, FKNN divides the fuzzy members of the sample into different groups according to the following equations [18].

$$u_i(x) = \frac{\sum_{j=1}^{k} u_{ij} \left(1/\|x - x_j\|^{2/(m-1)} \right)}{\sum_{j=1}^{k} \left(1/\|x - x_j\|^{2/(m-1)} \right)}$$

(2-12)

where $i = 1,2,\cdots,C$ and $j = 1,2,\cdots,k$ reflect the classes and k represents the number of nearest neighbors. Also, m is the length fuzzy component to identify the distance weight value to calculate the neighbor contribution compared to the member value. Finally, this value is chosen as $m \in (1,\infty)$. Also, $\|x - x_j\|$ is the distance between x and the jth nearest neighbor of x_j. The distance is often measured in terms of the Euclidean distance metric. u_{ij} is the degree of membership in the structure x_j of the training set of class i, among the k nearest neighbors of x.

There are two ways to represent u_{ij}. The first way is the definite membership, i.e. in each learning structure an example has full membership in a given class and no membership in other classes. The second way is the very limited fuzzy membership, i.e. the nearest neighbor of the training structure (x_k) is found and the membership of x_k in the class is possible as follows.

$$u_{ij}(x_k) = f(x) = \begin{cases} 0.5 + (n_j/k)^* 0.49, & if\ j = 1 \\ (n_j/k)^* 0.49 & ,\ if\ j \neq 1 \end{cases} \qquad (2\text{-}13)$$

n_j is the number of neighbors found in the jth class. It should be noted that the membership estimated by Relation 2 satisfies the following equations.

$$\sum_{I=1}^{C} \mu_{ij} = 1, \quad j = 1,2,\cdots,n \qquad (2\text{-}14)$$

$$0 < \sum_{I=1}^{n} u_{ij} < n,$$

$$u_{ij} \in [0,1]$$

After estimating the memberships for a query, the example is assigned to the class with the maximum membership value [18].

$$C(x) = \underset{i=1}{\overset{c}{argmax}}\big(u_i(x)\big) \qquad (2\text{-}15)$$

2-14 CRISP methodology

There are various methods for implementing and executing data mining projects. The CRISP method is one of the most powerful methodological methods, including the following steps.

- System identification
- Data identification

- Data preparation

- Modeling, evaluation, and system development

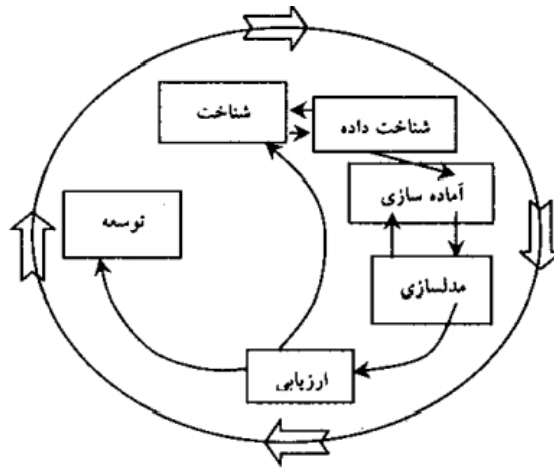Each step is divided into several sub-steps [14].

شناخت داده   شناخت

آماده سازی

توسعه

مدلسازی

ارزیابی

Figure 2- 4- Steps of the CRISP methodology [10]

2-15 Proposed method

As mentioned in the previous section, the proposed method of the basic paper [9] is divided into three parts, including image processing, feature extraction, and classification. One of the most important steps in this work is the classifier. The higher the learning power of the classifier, the higher the accuracy of the detection. For this purpose, we use a powerful CNN classifier to classify and diagnose the disease. Convolutional neural networks or CNNs are an example of deep neural networks that are usually used to perform image or speech analysis in machine learning. Convolutional neural networks use a type of multilayer perceptron to minimize preprocessing. Our goal is to increase the accuracy of Parkinson's disease diagnosis based on the learning ability and scalability of deep CNNs. These special capabilities provide significant advantages over the results of the reference paper.

Chapter Three: A Review of Resources

3-1 Literature Review

The movement of the hand while writing has been studied in [5]. The location and position of the pencil on the tablet were recorded and a vector of indicators such as the standard deviation of the speed, signal entropy was calculated to identify and represent the time and location of the pencil. Using the information and data of healthy individuals and PDs, different statistical indicators were presented between the two groups. Dopamine release in PD individuals is lower than normal. Also, ability to control the speed of movement of the parameter is affected. Changes in writing have multiple peaks. However, in healthy individuals, only one peak is seen.

Researchers in [3] studied the movement and function of the hand while writing. The movement of the hand during writing includes two components, namely the pressure exerted by the screen and the movement of the hand in the air while touching the screen. The study included 37 PD patients and 38 healthy individuals. The two groups were asked to write a specific sentence. They were provided with a normal-sized ink pen. All signals were recorded using a digital tablet. The number 0 was considered for the pencil being up (moving in the air) and the number 1 was considered for the pencil being down (moving on the page).
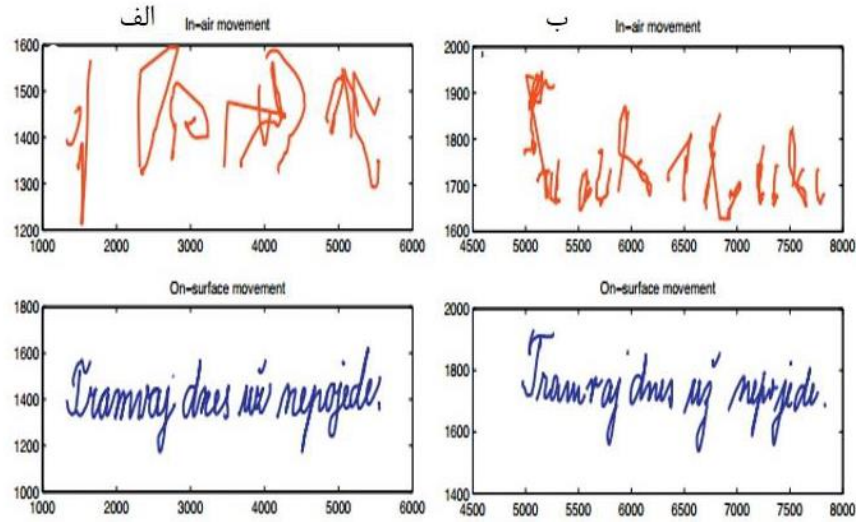
Figure 3- 1- is an example of hand movement on the surface (blue line) and in the air (red line) while writing [4]

Signal (a) corresponds to the healthy group and signal (b) represents people with Parkinson's disease. As it is known, the patient's hand movement in the air and on the paper has tremors.

In [6], the visual grapomotor feature of PD patients was evaluated. Neural networks were used to analyze the data. Therefore, different visual features were generated from different grapomotor samples. Finally, the extracted features were classified by a support vector machine and an accuracy of 83% was achieved based on visual information alone.

The researchers in [7] aimed to investigate the question "Do dynamic handwriting features have the ability to support the diagnosis of PD in the early stages?" For this purpose, a subset of the PaHaW dataset was used. Finally, a classification framework based on multiple different classes was presented. The movement disorder symptoms of advanced Parkinson's patients were measured using spiral data drawn by a tactile telemetry device.

31

In [8], movement symptoms were determined to automate the visual interpretation of the spiral data. Two groups of 65 Parkinson's patients and 14 healthy individuals were selected. Spatial and temporal features were used as input parameters of the vector machine. The best classifier (multilayer perceptron) was able to classify motor symptoms (slowness and involuntary repetitive movements) with an accuracy of 84% and an area under the curve of 86%. In clinical science, the main focus and goal is to determine the severity of symptoms of speech disorders, tremor, slowness, and abnormal repetitive movements. Analysis of the spiral drawn on the tablet is a sensitive and reliable solution to determine tremor in patients.

3-2 Basic paper

In [9], using computer vision-based techniques, the results of some methods on the automatic diagnosis of Parkinson's were obtained. The dataset used in this study was collected by clinical handwriting examinations of several clinics. The data was analyzed using image processing and machine learning techniques. In addition, a new numerical feature called the relative mean tremor is proposed to measure the degree of tremor of an individual's handwriting. The proposed method of the reference paper is divided into three parts: image processing, feature extraction, and classification.

Using computer vision-based classical image processing techniques such as combining a blurring filter and some morphological and thresholding operations, the handwriting track (HT) was extracted from the spiral template (ST). Then feature extraction was applied to the handwriting separated from the spiral template. Feature extraction was performed with the aim of describing ST and HT and then comparing them to evaluate the "difference" between the two images. These features are the root mean square (RMS) difference between the ST and HT radius, the maximum difference between the ST and HT radius, the minimum difference between the ST

and HT radius, the standard deviation of the difference between the ST and HT radius, the maximum HT, the minimum HT, and the standard deviation of the HT radius. Finally, the extracted features were classified using three classifiers called NB, OPF, SVM, and achieved an accuracy of 78.9%, 77.1%, and 75.8%, respectively.

Chapter Four: Proposed Method

4-1 Introduction

According to research conducted in the field of medicine, the handwriting of Parkinson's patients is different. Researchers want to obtain parameters to represent such changes. In the research conducted, different methods and solutions were examined and presented. The purpose of introducing and examining some solutions is to present a suitable proposal for processing handwriting for the diagnosis and identification of Parkinson's disease. In recognizing and examining handwriting patterns, the neural network classifier plays a vital role in classification. In this research, we use the convolution neural network (CNN) for data classification, which is described below.

In the rest of this chapter, we describe the database used and the evaluation criteria used. Finally, we introduce the steps of the proposed method of this study.

4-2 Database

One of the most challenging processes when facing the diagnosis of Parkinson's disease (PD) is the use of visual information and signals from patient examinations. Previous works used high-end imaging technology (MRI) for such purposes. However, this technology is expensive and may be dangerous for patients. In addition, most of the signal-based datasets for PD diagnosis are small and do not reflect the real aspects.

In order to overcome such shortcomings, we created a new dataset consisting of images extracted from handwritten examinations of 55 individuals. 37 individuals from the test group are included with Parkinson's disease (patient group) and the remaining 18 are included with healthy individuals (control group). Furthermore, the age of individuals in the patient group and control group varies within the range of [38, 78] and [19, 79], respectively. The patient group is divided into 25 males and 12 females. Meanwhile, the control group is included with 6 males and 12 females.

This dataset was collected at the Botucatu School of Medicine, State University of São Paulo, Brazil. The writing test basically consists of filling out a form in order to perform some tasks such as drawing circles, spirals, and vehicles. Figure 4-1 shows the exam of a 56-year-old male patient in which the essential tremor of Parkinson's disease is visible.
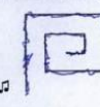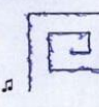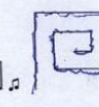
Figure 4- 1- Handwriting sample of a 56-year-old Parkinson's patient

Although we are faced with tasks related to spirals and curves, the first version of our dataset only includes images of spirals. Although we have 4 spirals in each exam, some subjects did not perform all of them. Therefore, the proposed dataset consists

of 373 samples (4 spiral samples per subject). Figure 4-2 represents examples of spirals extracted from the dataset.
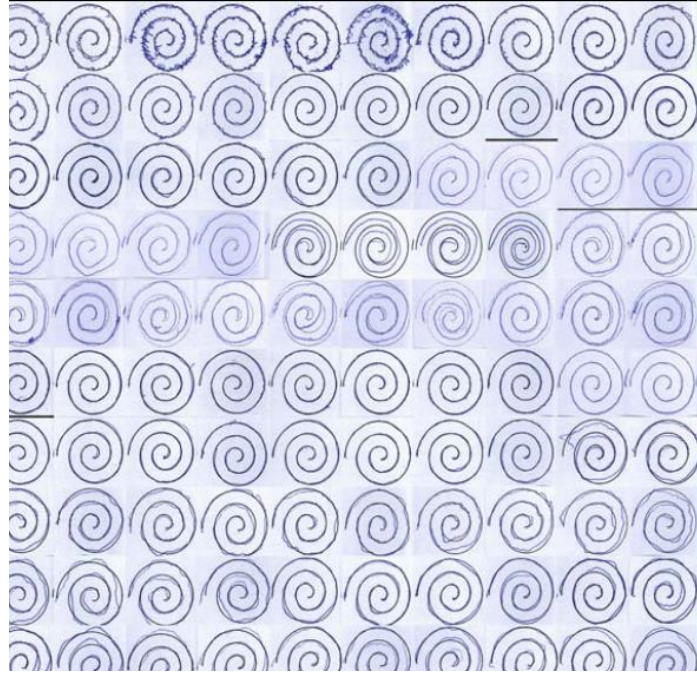
Figure 4- 2- Some samples of spirals extracted from the dataset

4-3 Evaluation criteria

Evaluation criteria are very important items to identify the most desirable technique for performing the test. In this study, the accuracy criterion is used and is defined as follows.

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

(4-1)

where TP is true positive, TN is true negative, FN is false negative, and FP is false positive.

4-4 Proposed method

In this study, handwriting images of people undergoing Parkinson's test are automatically examined by MATLAB software and patients and healthy people are separated by machine processing and machine learning methods. The MATLAB program consists of two parts, namely feature extraction and classification. In the feature extraction part, patients' handwriting images are fed. Then, handwriting and printed lines are separated and compared to obtain the relevant features. The angular profile of the image pixels is used to identify the corresponding points on the handwriting and printed lines. Then the features are generated in the same order as mentioned and are classified by the deep neural network in the next step. Figure 4-3 represents the block diagram of the proposed method.

فراخوانی تصاویر پایگاه داده

جدا کردن داده های دست نویس و چاپی

پردازش تصاویر

استخراج ویژگی ها از تصاویر

تشکیل ماتریس داده و هدف

نرمال سازی داده ها
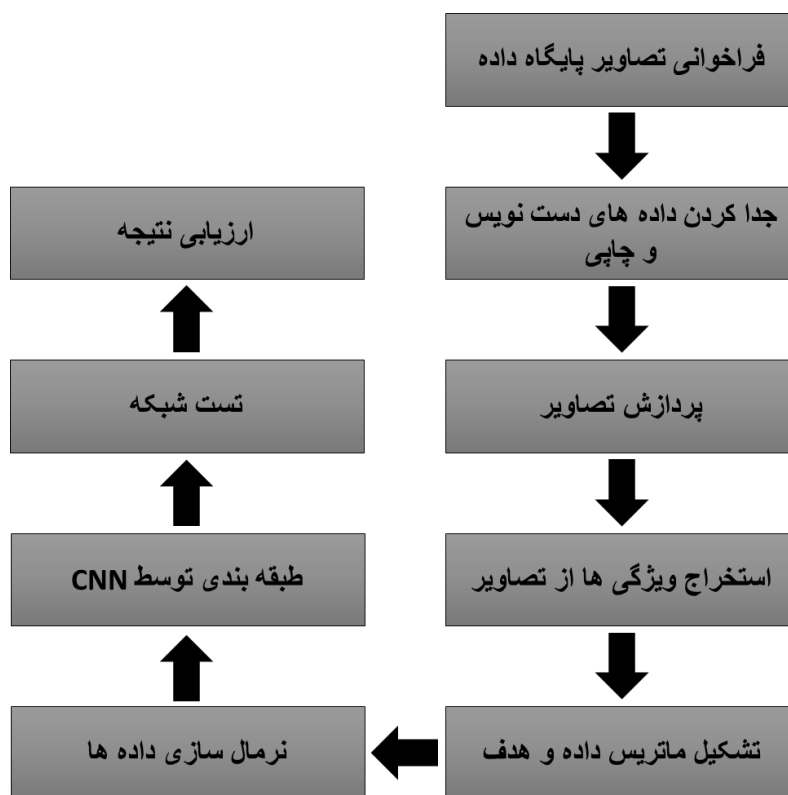
طبقه بندی توسط CNN

تست شبکه

ارزیابی نتیجه

Figure 4- 3- Block diagram of the proposed method

In this section, we describe the proposed dataset design method of handwriting test. To achieve this, we divide the proposed method into two parts, namely (i) image processing and (ii) feature extraction. First, we need to extract the handwriting trace (HT) from the spiral template (ST). This is because the images are intertwined. Later, we use HT and ST to calculate features based on their shapes. Figure 4-4 shows an example of the input image and the corresponding spiral template and handwriting.
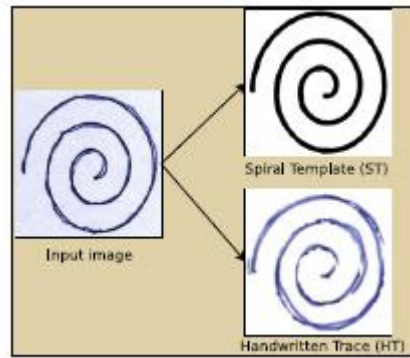
Figure 4- 4- A sample of the input image and the corresponding spiral template and handwriting

4-4-1 Image processing

In order to segment ST and HT, we combined some classical image processing techniques such as blurring filters and mathematical morphology. The extraction processes of both ST and HT contours are performed separately. Since the images are digital, we performed a preprocessing step to reduce noise and unwanted artifacts using a 5×5 mean filter. Then, by thresholding the smoothed image, the spiral template was extracted to obtain a binary M_{ST}^{i} mask of the spiral. This process is performed as follows.

$$M_{ST}^{i} = \begin{cases} 0 & \text{if } R^i < 100 \wedge G^i <| 100 \wedge B^i < 100 \\ 1 & \text{otherwise,} \end{cases}$$

(4-2)

where R^i, G^i, and B^i are the values of the ith pixel of the input image in the "Red", "Green" and "Blue" channels respectively. In short, if Equation 4-2 holds, the foreground (spiral template) pixels are set to 1 ("black") and the background pixels are set to "white" as shown in Figure 4-4. Suppose the pixel in the original (input) image is black or almost black. It is reasonable to assume low brightness values for such pixels when searching for the spiral itself. Finally, to ensure a fully-connected spiral template, we used an open operation (erosion followed by dilation). Figure 4-5 represents the proposed path for ST extraction.
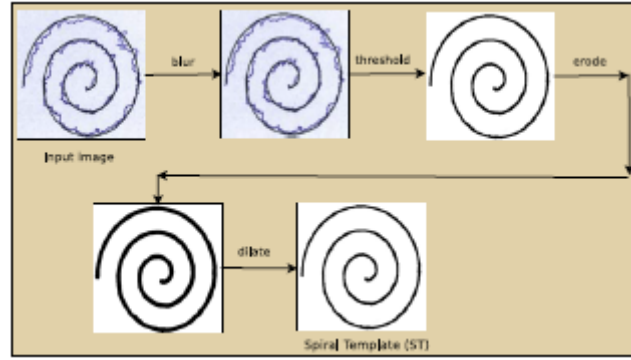
Figure 4- 5- Image processing steps in the spiral template extraction process

In the HT extraction step, we used a similar approach to the previous method used to extract the spiral pattern. However, some additional steps and a different thresholding method are applied, since the HT and the background are blue. In the first step, we used mainly HT edge filters, 5×5 median filter, followed by 5×5 median filter to smooth the image to reduce noise and small artifacts. Furthermore, thresholding on the filtered image F is performed using the following equation.

$$M_{HT}^i = \begin{cases} 255 & \text{if } |R^i - G^i| < 40 \wedge |R^i - B^i| < 40 \wedge \\ & \wedge |G^i - B^i| < 40 \\ F^i & \text{otherwise,} \end{cases}$$

(4-3)

where F^i stands for the brightness of the ith pixel. The visual idea behind this step is to remove pixels with pseudo-similar values for the three channels (i.e., background pixels). Also, pixels with significant differences between the channels (background - HT - pixels) are retained. Figure 4-6 reflects the proposed path for HT extraction.
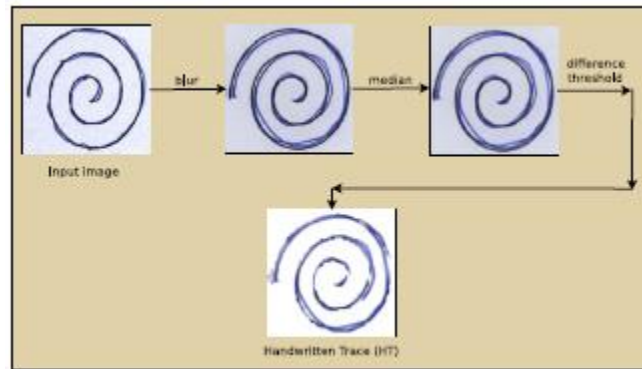
Figure 4- 6- Image processing steps in the case of manual tracking

4-4-2 Feature extraction

The objective of feature extraction is to describe ST and HT and then compare them to assess the "difference amount" between the two images. To achieve this goal, we need a concise and compact representation of ST and HT. This process is performed here using a skeleton of thresholded images. Therefore, we extracted the skeletons of ST and HT images based on the thinning algorithm proposed by Zhang Suen [17]. This algorithm consists of two parallel routines including (a) removing the southeast boundary points and northwest corner points and (b) removing the northwest boundary points and southeast corner points. Figures 4-7 and 4-8 show the thinning results of the spiral template and the handwriting path, respectively.
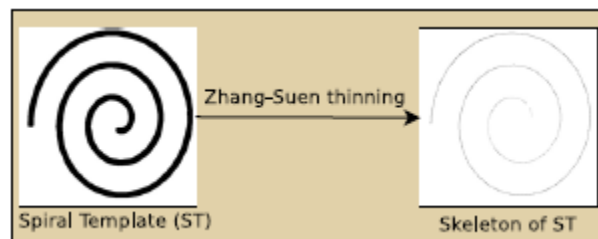
Figure 4- 7- Thinning of the spiral template using Zhang Suen's algorithm
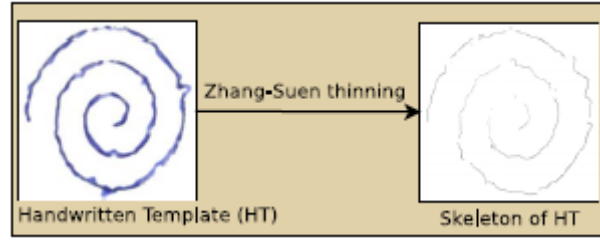
Figure 4- 8- Thinning of the handwriting path using Zhang Suen's algorithm

Then, we extracted 9 numerical features from each skeleton (i.e., ST and HT) by measuring the statistical differences between them. Before describing the feature, we introduce the "radius" of a spiral point. The radius is actually the length of the straight line connecting a given point to the center of the spiral, as shown in Figure 4-9. The "red" point means the center of the spiral. Also, some random points ("white") are connected to the thin spiral line (skeleton).

Figure 4- 9- Some random points and a straight line connecting the spiral center point

A brief description of each feature is given below.

1) Root mean square (RMS) is based on the difference between the ST and HT radii. RMS is calculated as follows.

$$\text{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_{ST}^i - r_{HT}^i)^2},$$

(4-4)

where n is the number of sample points extracted from each ST and HT skeleton. Also, r_{ST}^i and r_{HT}^i represent the ith sample point of the ST and HT radii, respectively.

2) Maximum difference between ST and HT radii:

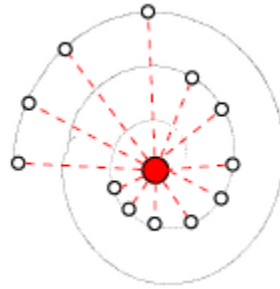$$\Delta_{max} = \underset{i}{\text{argmax}}\{|r_{ST}^i - r_{HT}^i|\};$$

(4-5)

3) Minimum difference between ST and HT radii:

$$\Delta_{min} = \underset{i}{\text{argmin}}\{|r_{ST}^i - r_{HT}^i|\};$$

(4-6)

4) Standard deviation of the difference between ST and HT radii.

5) Mean relative jitter (MRT): We proposed this numerical evaluation technique to measure the "jitter amount" of an individual's handwriting. This metric is defined as the average difference between the radius of a given sample and its d nearest left neighbors. MRT is calculated as follows:

$$\text{MRT} = \frac{1}{n| - d} \sum_{i=d}^{n} |r_{HT}^i - r_{HT}^{i-d+1}|,$$

(4-7)

where n is the number of sample points and d is the amount of displacement of sample points to calculate the radius difference. The following three features are calculated based on the relative jitter.

6) Maximum HT.

7) Minimum HT.

8) Standard deviation of HT values.

9) Number of times the difference between ST and HT radius changes from negative to positive or vice versa.

4-4-3 Classification by CNN deep neural network
After extracting the image features, the data must be classified by an appropriate classifier. In this study, an 8-layer convolutional neural network was used. Using fewer layers or a different algorithm may affect the behavior of the neural network. The components of this 8-layer structure are:

- Input layer
- Convolutional layer
- ReLU layer
- First fully-connected layer
- Second fully-connected layer
- Third fully-connected layer
- Softmax layer
- Classification layer

Below, we briefly describe these layers.

4-4-3-1 Input layer
Images are the main input to the network. Next, the information is normalized.

4-4-3-2 Convolution layer
By passing a filter along the input in the vertical and horizontal directions, the point output of the weight and input are estimated, and then a bias is added.

4-4-3-3 ReLU layer
Rectified linear unit performs the threshold function for each of the input parameters. Values less than zero represent zero.

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$(4\text{-}8)$$

4-4-3-4 Fully-connected layer

The input parameter of the fully-connected layer is multiplied by the weight matrix and the bias vector is added.

4-4-3-5 Softmax layer

The Softmax function is applied to the input.

4-4-3-6 Classification layer

The cross-entropy loss is estimated for multi-class classification problems with a special and unique class. This layer receives its input from the output of the previous layer. To identify the k classes of the network, the fully-connected layer requires the output value of k and the softmax layer before the classification layer.

Chapter Five: Results and Suggestions

5-1 Introduction

As mentioned in the previous section, the proposed method of the basic paper [9] is divided into three parts, including image processing, feature extraction, and classification. One of the most important steps in this work is the classifier. The greater the learning power of the classifier, the higher the detection accuracy. For this purpose, we use the powerful CNN classifier for disease classification and diagnosis. Convolutional neural networks (CNNs) are a type of deep neural network that is commonly used to perform image or speech analysis in machine learning. Convolutional neural networks use a type of multilayer perceptron to minimize preprocessing. Our goal is to improve the accuracy of Parkinson's disease diagnosis based on the learning ability and scalability of deep CNNs. These special capabilities provide significant advantages over the results of the reference paper.

In the following sections of this chapter, the database used is presented. Then, the data normalization operation is described, and finally the classification results are compared with previous works.

5-2 Database

In this simulation, we considered two sets of data in two separate folders. 296 images are related to the patient group. Since each person provided 4 images (some also provided 3 images), in fact, the data is related to 74 people. The remaining 72 images belong to 18 healthy people. In total, there are 92 patient and healthy individuals in this database. After processing the images and extracting the features, the dimensions of the data matrix will be 92 by 9. Figure 5-1 shows a sample of this data in Matlab.

	1	2	3	4	5	6	7	8	9
79	19.6122	131.1338	0.5033	26.9106	22.6063	348.7170	37.8987	84.5718	28
80	24.9328	139.5352	0.1651	35.3234	34.0080	345.6430	48.4557	78.0448	21.5000
81	23.2055	128.1254	0.1640	35.9361	32.7700	344.0801	47.9628	80.0688	19.5000
82	19.6215	136.1000	0.0880	33.2561	31.4553	335.6273	43.7316	87.3632	23
83	11.6383	121.2866	0	20.5979	22.3180	347.8027	50.5969	89.8180	45
84	22.7155	145.1017	0.9025	26.5855	26.9750	342.2414	51.3501	81.4006	19.5000
85	16.7224	162.5065	0.4280	25.6458	23.7051	348.5506	42.2189	90.2986	20
86	7.1389	98.3003	0.0401	12.1622	14.8319	351.4402	48.9996	91.9611	31.5000
87	9.5231	93.9178	2.7270e-04	12.3956	14.5469	340.0623	44.9471	89.9279	24.5000
88	9.7417	126.7784	0.0985	19.2270	17.6749	353.6732	47.0117	92.6198	35.5000
89	14.4039	119.2552	0.4395	14.6599	13.6824	341.0684	33.5298	93.5704	20.5000
90	18.7558	122.2348	0.2657	28.8399	24.9522	336.6995	48.2482	81.2943	27.5000
91	7.2373	124.6505	0	17.0345	17.3692	335.2993	55.9164	90.1249	55
92	8.6710	126.9659	0.0391	18.4985	19.5054	344.2694	46.9511	91.5816	34.5000

92x9 double

Figure 5- 1- Sample data fed to the classifier

Also, our target matrix indicates whether the individual is healthy or patient and is represented as a 92 by 1 matrix. In this matrix, the number 1 indicates that the individual is healthy and the number 0 indicates that the individual is patient.

5-3 Data normalization

At this stage, the data must be normalized in the interval [-1,+1] to enter the network. The following relation is used for this transformation.

$$X_n = 2\frac{X_i - X_{min}}{X_{max} - X_{min}} - 1,$$
$$i = 1,2, \dots, N \tag{5-1}$$

where X_i is the actual value of the network input and X_n provides the normalized value. X_{min} and X_{max} are the minimum and maximum values of X, respectively.

After preparing the data and determining its status, the database is formed and the classification process enters the next step.

5-4 Network training

After normalization, the data is divided into two parts: test and training data. In this work, 80% of the data is considered as training data and 20% for testing. Figure 5-2 shows the characteristics of the CNN neural network after training. The time required to train the network was 19 seconds and lasted 100 epochs.

Results

Validation accuracy:	N/A
Training finished:	Reached final iteration

Training Time

Start time:	01-Aug-2020 19:39:56
Elapsed time:	19 sec

Training Cycle

Epoch:	100 of 100
Iteration:	100 of 100
Iterations per epoch:	1
Maximum iterations:	100

Validation

Frequency:	N/A
Patience:	N/A

Other Information

Hardware resource:	Single GPU
Learning rate schedule:	Constant
Learning rate:	0.01

Figure 5- 2- Characteristics of the CNN neural network after training

5-5 Simulation and comparison results

Figure 5-3 presents the classification results of the proposed neural network.
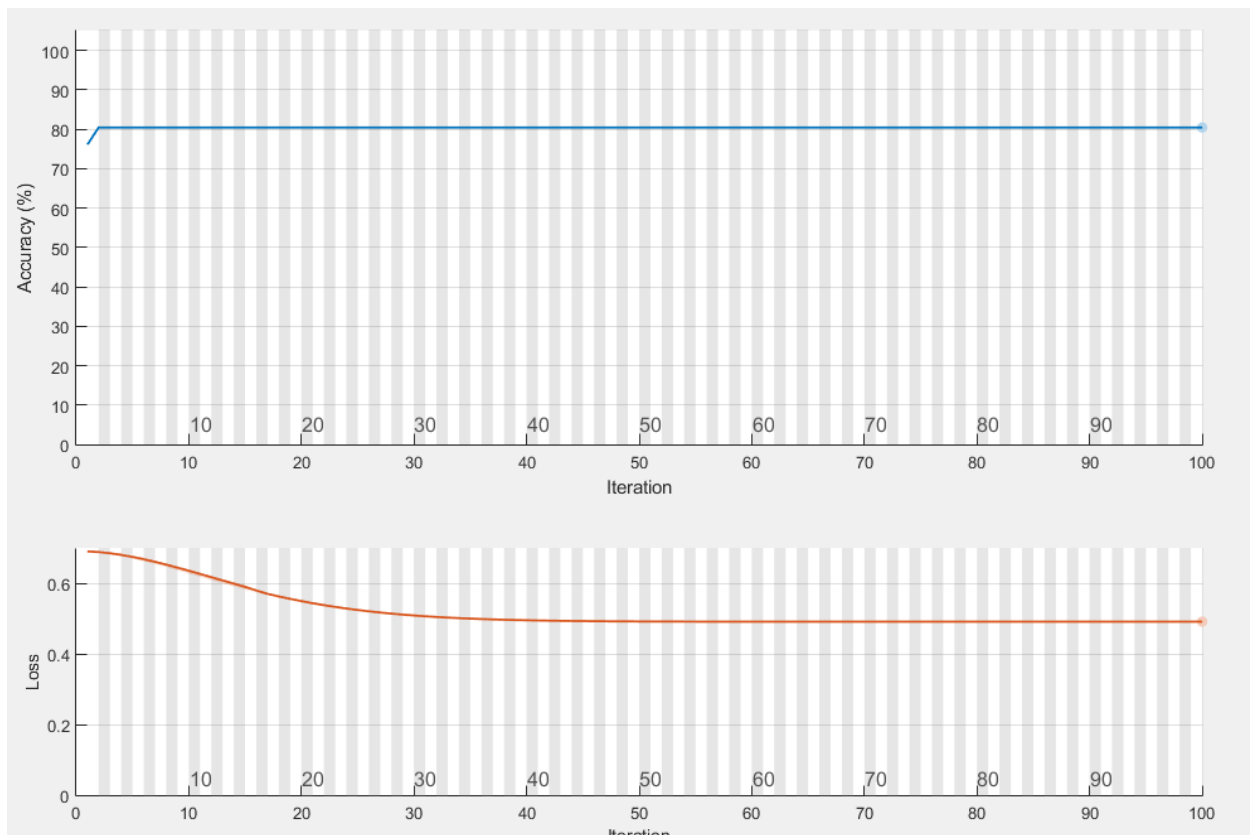
Figure 5- 3- Accuracy plot of the deep neural network

As it is clear, our network converged quickly and reached an accuracy of about 80%.

Table 5-1 compares the accuracy of different classifiers with the method presented in this research.

Table 5- 1- Comparison of results of different methods

CNN	SVM	OPF	NB	METHOD
80.4%	75.8%	77.1%	78.9%	ACCURACY

As it is clear, the CNN classifier performs better than other classifiers and achieves higher accuracy. The reason is that deep neural networks use less preprocessing than other image classification approaches, that is, the network learns parameters that were learned manually in previous approaches.

5-6 Recommendations

The following are recommended to increase the classification accuracy in future work.

1. Use new convolutional neural networks such as Resnet, googlenet, etc. for classification.

2. Use feature selection algorithms to select efficient features.

3. Use different databases to accurately measure the accuracy of the proposed algorithm

4. Use Resnet, googlenet convolutional networks to extract efficient features

References

[1] C. Pereira, D. Pereira, F. Silva, J. Masieiro, S. Weber, C. Hook, J. Papa, "A new computer vision-based approach to aid the diagnosis of Parkinson's disease", computer methods and programs in biomedicine, vol. 136, p. 79-88, 2016.

[2] E. Smits, A. Tolonen, L. Cluitmans, M. Gils, B. Conway, R. Zietsma, K. Leenders and N. Maurits, "Standardized Handwriting to Assess Bradykinesia, Micrographia and Tremor in Parkinson's Disease", Plos one 9, vol. 5, 2014

[3] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, M. Faundez-Zanuy, "Analysis of in-air movement in handwriting: A novel marker for Parkinson's disease", computer methods and programs in biomedicine, vol. 3848, p. 7, 2014.

[4] S. Rosenblum, M. Samuel, S. Zlotnik, I. Erikh, I. Schlesinger, "Handwriting as an objective tool for Parkinson's disease diagnosis", J Neurol, vol. 260, p. 2357-2361, 2013.

[5] C. Kotsavasiloglou, N. Kostikis, D. Hristu-Varsakelis, M. Arnaoutoglou, "Machine learning-based classification of simple drawing movements in Parkinson's disease", Biomedical Signal Processing and Control, vol. 31, p. 174-180, 2016.

[6] Moetesum, M., Siddiqi, I., Vincent, N., & Cloppet, F. (2019). Assessing visual attributes of handwriting for prediction of neurological disorders—A case study on Parkinson's disease. Pattern Recognition Letters, 121, 19-27.

[7] Impedovo, D., Pirlo, G., & Vessio, G. (2018). Dynamic handwriting analysis for supporting earlier Parkinson's disease diagnosis. Information, 9(10), 247.

[8] Pereira, Clayton R., et al. "Deep Learning-Aided Parkinson's Disease Diagnosis from Handwritten Dynamics." Graphics, Patterns and Images (SIBGRAPI), 2016 29th SIBGRAPI Conference on. IEEE, 2016.

[9] Pereira, C. R., Pereira, D. R., da Silva, F. A., Hook, C., Weber, S. A., Pereira, L. A., & Papa, J. P. (2015, June). A step towards the automated diagnosis of parkinson's disease: Analyzing handwriting movements. In 2015 IEEE 28th international symposium on computer-based medical systems (pp. 171-176). IEEE.

[10] Keshavarz, Hamd, Ghasemian Yazdi, Hassan, 2005, A fast algorithm based on support vector machines for classifying hyperspectral images using spatial correlation, Iranian Electrical and Computer Engineering 1.37. Mortezapour, Reza, Motlaghzadeh, Mahsan,

2012, Using Bayesian networks in classification, Fourth Iranian Electrical and Electronics Engineering Conference

[11] Arefian, Fatemeh, Eftekhari, Mahdi, 2013, A new fuzzy and uneven K-nearest neighbor method for semi-supervised classification, National Conference on Computer Engineering and Sustainable Development with a focus on computer networks, modeling and systems security.

[12] Gholami, Mohsen and Najafi, Naser, 2016, Study and comparison of the effectiveness of data mining algorithms for predicting Parkinson's disease, First International Conference on New Perspectives in Electrical and Computer Engineering, Tehran, The International Federation of Inventors' Associations (IFIA), University of Applied Sciences and technology.

[13] Tahmasebi, Hamid Reza and Amuzgar, Malihe, 2012, Application of Data Mining in Medicine, Fourth Iranian Conference on Electrical and Electronic Engineering, Gonabad, Islamic Azad University, Gonabad Branch.

[14] Ghazanfari, Mahdi, Alizadeh, Somayeh and Teymourpour, Babak, 2008 Data Mining and Knowledge Discovery, Iranian Science and Industry Publishing Center.

[15] Xiaowei Han, Michline Camber, Jean Pi, 2015 Data Mining: Concepts and Techniques, translated by Mahdi Esmaeili, Niaz-e Danesh Publishment.

[16] Toloei Ashlaghi, Abbas, Pourabrahimi, Ali, Ebrahimi, Mandana, Ghasem Ahmad, Leila, 2012, Prediction of breast cancer recurrence with three data mining techniques. Iranian Quarterly Journal of Breast Diseases, Volume 5, Number 4.

[17] Ghermezkon, Mahdi, Mousavi Rad, Seyed Jalaleddin, Ghermezkon, Mohammad, 2015 Improving Parkinson's disease diagnosis using AdaBoost algorithm, International Conference on New Research Findings in Electrical Engineering and Computer Science, Tehran, Nikan Institute of Higher Education.

[18] Hui-Ling Chen, et al. (2013). An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. Expert Systems with Applications, 40, 263-271.