Classification of Alzheimer Disease based on Normalized Hu Moment Invariants and Multiclassifier

Arwa Mohammed Taqi
Systems Engineering Department
University of Arkansas at Little Rock
Little Rock, AR 72204, USA

Fadwa Al-Azzo
Systems Engineering Department
University of Arkansas at Little Rock
Little Rock, AR 72204, USA

Mariofanna Milanova
Computer Science Department
University of Arkansas at Little Rock
Little Rock, AR 72204, USA

Abstract—There is a great benefit of Alzheimer disease (AD) classification for health care application. AD is the most common form of dementia. This paper presents a new methodology of invariant interest point descriptor for Alzheimer disease classification. The descriptor depends on the normalized Hu Moment Invariants (NHMI). The proposed approach deals with raw Magnetic Resonance Imaging (MRI) of Alzheimer disease. Seven Hu moments are computed for extracting images' features. These moments are then normalized giving new more powerful features that highly improve the classification system performance. The moments are invariant which is the robustness point of Hu moments algorithm to extract features. The classification process is implemented using two different classifiers, K-Nearest Neighbors algorithm (KNN) and Linear Support Vector Machines (SVM). A comparison among their performances is investigated. The results are evaluated on Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The best classification accuracy is 91.4% for KNN classifier and 100% for SVM classifier.

Keywords—Alzheimer disease; machine learning; Hu moment invariants; SVM; K-Nearest Neighbors (KNN) classifier

I. INTRODUCTION

Alzheimer's disease (AD) is a permanent, progressive neurological brain disorder and complex disease which gradually destroys brain cells, reducing memory and thinking ability causing the dead, and eventually loss of the capability to perform even the simplest tasks. The mental weakening produced by this illness leads to dementia in the end [1]. AD was named after the German psychoanalyst and pathologist Alois Alzheimer when he tested a female patient (post mortem) in 1906 [2]. The first area affected is the hippocampus, which is responsible for episodic and spatial memory and works as a communicate structure between the brain and the body. The hippocampus shrinks unusually in an AD patient; where the normal decrease is between 0.24 and 1.73 percent yearly, a hippocampus imposed with AD might shrink between 2.2 and 5.9 percent [3].

Magnetic resonance imaging (MRI) is a medical imaging technique used in radiology to examine the body's anatomy and physiology in both healthy and diseased patients [4]. MRI scans can offer a utilitarian tool for estimating the properties of anti-dementia drugs in clinical tests which can highly serve the researchers in these fields. Scans can provide information about the levels and location of cell damage over time, and that would help to get valuable information about the optimistic effects of potential treatments.

Basically, the most distinctive structure which different a normal brain from a pathological brain is the symmetry. If it is clear from any view in either coronal or axial directions, that indicates a normal brain, and if broken it is a pathological brain [5]. Though, sometimes there is a strong similarity between the normal cognitive brain image and the brain with AD, especially when the brain is compromised by the start of AD. These situations also result in distinguishing the disease correctly. To beat this problem and to enhance the recognition accuracy, Hu moments approach presented by Hu [6] is used in this work, where the values are invariant with respect to the scale, translation, and rotation. Moment invariants were chosen because they are one of the most important and most used methods in the object recognition field.

In this work, seven-moment invariants have been computed for each image of normal cognitive and AD cases, and they are kept as a 1D vector. These calculations are repeated for the testing dataset, too. To highly improve the classification performance, these moments are then normalized to get more efficient features, which can be easily distinguished by the classifiers later. Two different classifiers are used for the classification process, KNN and SVM, to measure the minimum matching between the training and testing datasets for each class. Minimum matching indicates the closer case of testing dataset to the specific class of the training dataset.

The organization of this paper is as follows: Section II illustrates the related work with this study. Section III explains the Hu moments theory. Section IV describes the proposed material and methods. The NHMI is trained based on that database by applying Hu moment invariants algorithm designed for feature extraction. Section V explains a comparison between KNN and SVM classifiers. Section VI shows the experiments' results. Finally, Section VII and VIII summarizes the conclusions and discussion of this work.

II. Literature Review

Numerous systems have been used to solve classification problems. One of the influential methods is machine learning algorithms. Hu moments theory has been considered as a powerful way to extract the dominant features of an image as we did in our earlier work in [7]. It highly proved the strength of the extracted features for different human actions. On other hand, Support Vector Machine (SVM), an influential binary classifier, is one of the most broadly used classifiers. It is suitable for high dimensional classification problems, where not too many examples exist. SVM has been utilized, for instance, in [8] for classifying MR images and in [9], [10] to classify Position Emission Tomography (PET) images. All these works used voxel intensity (VI) as features. In a different approach, a single multi-kernel SVM has been employed for the multimodal classification of MRI, PET, and CSF using VI within regions of interest [11]. Although SVM has been the preferred single classifier, other options such as Gaussian Naives Bayes [4], [12] or Gaussian Processes [13] have also been used successfully.

Another way to classify accurately is to use ensembles which combine the outputs of several classifiers. Several well-known ensemble methods have already been explored for AD classification. For instance, [14] propose the favorite class ensemble of classifiers where each base classifier in the ensemble uses a different feature subset which is optimized for a given class. In [15], an ensemble classifier was learned from different random subsets of local patches. Ensemble methods have also been used in order to combine information from different modalities such as EEG, MRI and PET [16]. Many of these methods use a prior feature selection step in order to reduce dimensionality. Different techniques have been used for this purpose, such as PCA [9] or selecting the best ranking features according to some criteria such as the t-test [11].

An Atrophy Differential Diagnosis Approach for early detection of Alzheimer disease (AD), where the atrophy is located on the brain and it offers hippocampus, a regional atrophy analysis for differential diagnosis of different neurodegenerative diseases, which is a computer aided system [17]. Wavelet Fuzzy C- Means (WFCM) algorithm is used for image segmentation in noisy medical images. The feature extraction is done by wavelet decomposition, and the feature vector is fed as input to FCM [18].

Another method to classify AD is to use voxel-wise, cortical thickness, and hippocampus shape volume features of the sMRI [19]. In this method, the first step is co-aligning (registering) all the brain image. So, each brain voxel will be associated with a vector of many scalar measurements. Then, voxel-wise features were extracted. While the work of [20] extracted the features by using the gray matter (GM) voxels, and use them to train an SVM to distinguish between the AD and NC subjects. A different view to extract features is from brain volume. The work of [21] was segmented the brain volume to GM, white matter (WM), and CSF parts, and then estimate all voxel-wise densities and relating each voxel with a vector of GM, WM, and CSF densities for classification.

III. Hu Moments Theory

The moment invariants were initially presented by Hu [6]. Hu moments algorithm is selected to extract image features since the created features are rotation scale translation. Basically, Geometric Moment (GM) was effectively used in aircraft documentation, texture classification, and radar images for optical images matching [22].

Basic terms in the construction of the invariant moments have two steps. First, consider an image that has a gray function $f(x, y)$ having a bounded support and a finite nonzero integral. Second, geometric moment m_{pq} of the digital sampled $M \times M$ image [$f(x, y)$] which can be computed using (1) [23].

$$m_{pq} = \sum_{x0}^{x=M-1} \sum_{y=0}^{y=M-1} (x)^p \cdot (y)^q f(x, y),$$ (1)

p,q = 0,1,2,3,....., where p,q are non-negative integers and (p+q) is called the order of the moment.

The moments of f (x, y) are translated by an amount (a, b), which is calculated by (2).

$$\mu_{pq} = \sum_x \sum_y (x+a)^p \cdot (y+b)^q f(x, y)$$ (2)

Consequently, the central moment μpq can be calculated from (2) by replacing $a = -\bar{x}$, and $b = -\bar{y}$ as

$$\bar{x} = \frac{m_{1,0}}{m_{0,0}}, \bar{y} = \frac{m_{0,1}}{m_{0,0}},$$

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p \cdot (y-\bar{y})^q f(x, y),$$ (3)

The central moment of the image is invariant to translation, while the scaling invariance can be achieved by normalizing the moments of the scaled image by the scaled energy of the original image that can be computed as stated below:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu^{\gamma}_{00}}, \gamma = \frac{p+q}{2} + 1,$$

where γ is the normalization factor.

$$M_1 = \eta_{20} + \eta_{02},$$
$$M_2 = (\eta_{20} - \eta_{02})^2 + 4\eta^2_{11},$$
$$M_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2,$$
$$M_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2,$$
$$M_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[(\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2],$$
$$M_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}),$$
$$M_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})] + (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} - \eta_{03})^2].$$ (4)

In fact, Hu defined seven values, calculated by normalizing central moments completed order three that are

invariant to object scale, position, and orientation. In terms of the central moments, the seven moments are given as shown in (4) [24].

IV. MATERIAL AND METHODOLOGY

The proposed NHMI approach goals to obtain the more powerful features of the brain images for both healthy and AD cases. It extracts the features of the training and testing datasets using HMI algorithm. The extracted features of each image for both training and testing datasets are then normalized, representing the distinctive features of that image which results in a better classification performance. Subsequently, the classification process is taken over using two different supervised classifiers; KNN, and Linear SVM. Eventually, the closed features of the maximum matching would be selected as a matching output class. Fig. 1 shows the block diagram of NHMI approach.

A. Data

The investigated data in this work was obtained from the ADNI (Alzheimer's Disease Neuroimaging Initiative) database http://www.adni-info.org/. The ADNI initiative includes a longitudinal multi-modal track of all applicants through 36 months in which bio specimen, imageology, and clinical data were composed.

ADNI began its work in 2004. It is an enormous, 7-year effort to support and assist the discovery and development research that limits or restricts the growth of AD. Its target is to govern the features of AD as the pathology which grows from normal cognitive to mild symptoms, to Mild Cognitive Impairment MCI, and finally to dementia. ADNI is dedicated to creating standardized methods for imaging/biomarker groups and analysis to be used in clinical trials. In this paper, MRI core is only the interested core [25].

Generally, ADNI's subjects are between 55-90 years old, of both genders male and female. They have a study partner that can offer an independent estimation of functioning. Basically, there are two important criteria for diagnosing AD; Mini-Mental State Examination (MMSE), and Clinical Dementia Rating (CDR). MMSE ranges between 0-30, while CDR has five values; 0, 0.5, 1, 2, 3. For healthy applicants; MMSE scores are between 24-30, CDR of 0, this case refers to non-depressed, non-MCI, and non-demented subjects. With MCI subjects, MMSE scores between 24-30, but they have objective memory loss and a CDR of 0.5, basically conserved activities of daily life with absence of dementia. They do not classify as AD. Nevertheless, if MMSE scores are less than 20, and CDR scores are more than 0.5 (1, 2, and 3), that case would be considered as AD. They illustrate measures of disease severity.

In this paper, 100 subjects have been selected for training purposes, 50 with healthy controls (normal cognitive or MCI), and 50 subjects with AD. While another 28 subjects are used for testing purposes, 16 of them with AD and 12 of healthy subjects (normal control). An example of the used data in this work are shown in Fig. 2.

B. Feature Extraction and Selection

Features extraction process is a technique of image conversions, which transfers high-dimension features to the low-dimension features vector. In other words, the feature extraction achieves dimensional reduction at the same time it preserves the valuable information, which is most representative and essential to the image [26]. Features selection is an outstanding process among the most significant steps in image recognition, which could highly influence upcoming recognition process phases [27].
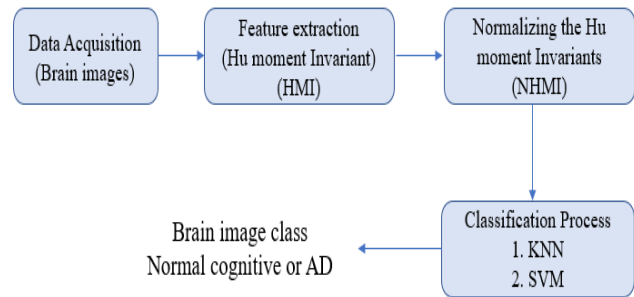

Fig. 1. Block diagram of the proposed technique.

TABLE I. HU MOMENTS FOR FOUR DIFFERENT BRAIN IMAGES

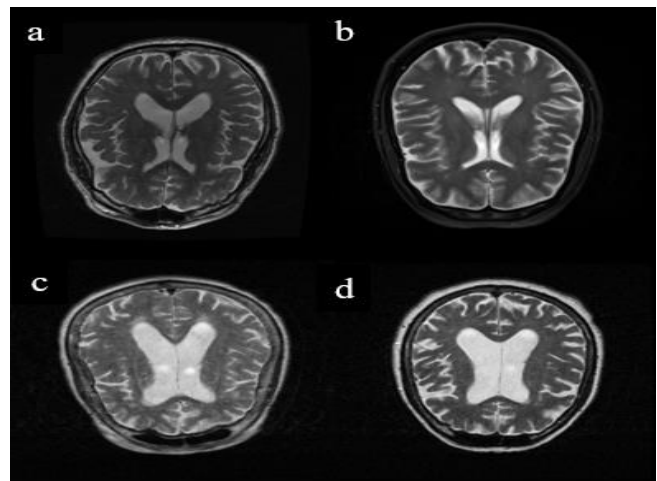| Hu moments	Image1 (CDR=0)	Image2 (CDR=1)	Image3 (CDR=2)	Image4 (CDR=3)
M1	1.5176	1.7704	1.6916	1.9289
M2	0.1508	0.6493	0.4815	0.9766
M3	0.7273	0.4747	0.2239	0.9659
M4	0.5079	0.2675	0.0367	0.7407
M5	0.1245	0.6307	0.1517	0.6169
M6	0.0833	0.0922	0.7775	0.7290
M7	-8.8941	10.4012 -	9.9237-	-11.4032

Fig. 2. MRI brain images: (a) Healthy brain with CDR=0. (b) AD brain with CDR=1. (c) AD brain with CDR=2. (d) Severe AD brain with CDR=3.

It is obvious that HMI is a set of seven invariants moments which can be used in applications that require scale, translation and rotation invariants. Truly, in this work, feature extraction process contains calculating seven Hu moments for each brain image as in (4), and all moment's values are concatenated into a 1D vector. Under those circumstances, a vector of seven Hu moments has been calculated for each brain image for both cases the normal control and AD. Therefore, each 2D brain image is transformed to a 1D vector containing the most significant feature of that image. Table 1 shows an example of seven Hu moments for four different brain images as in Fig. 2.

C. Normalizing the Hu moments

As it is clear from Table 1, there is some convergence among moments' values, which make it confuse for the classifiers to make the classification decision. To handle this issue, normalizing these moments has been found as the perfect way to diverge among them and make them specific features for each category to which they belong. The input features are normalized to real values between 0 and 1. The normalized moments have dissimilar values.

Normalization is the method used to reduce the needless repetition of data i.e., redundant data. It makes the data in a normalized arrangement. The main advantage of normalizing is to separate data into distinct, unique sets. Mostly, data normalizing is performed to improve the performance. Database normalization is a sequence of steps followed to get a database structure that permits reliable storage and effective access of data in a relational database. These steps decrease data redundancy and the hazard of data being unpredictable. Normalizing a database helps design the database construction to store data in a rational and related way. It is common for all databases to be normalized. First, normalizing data could reduce data duplication. Since databases can hold a significant amount of information, maybe millions or billions of pieces of data, normalizing the database reduces its size and prevents data duplication from happening. It makes sure that every piece of data is stored just once. Second, normalizing can group data logically.

Practically, application providers who make applications dealing directly with the database discover it is easier to treat with a normalized database. The data is arranged more logically when it is normalized. Normalizing gives fewer null values and less redundant data, making the database more compact. Conceptually, normalization is cleaner and easier to preserve and change whenever change is needed. As a result, normalizing highly improves the performance of the two classifiers used for identifying each image's class exactly. Table 2 displays the normalized Hu moments that shown in Table 1. Fig. 3 illustrates the feature extraction process.

D. Classification Process

The similarity measurement among images is still a hot topic and an essential issue in the machine learning and computer vision. Several applications in machine learning have usually used the Euclidean distance, for example, K-Nearest Neighbor (KNN), K-Means Clustering (KMC), and the Gaussian kernel. Some of them have used a binary classifier like SVM. Each classifier has specific characteristics, in terms of time consumption, performance accuracy, and cost, which make it the proper classsifier for some applications.

The basic classifiers that are used in this study are KNN and linear SVM. The main idea of using different classifiers is to demonstrate that the proposed technique is appropriate for more than one classifier. The option of these classifiers rested in KNN and SVM, both are suitable to high dimensional application, particularly when the available training examples are quite few. Both KNN and SVM are distinctive classifiers; they attempt to estimate classification limits in the feature space as a substitute of modelling the conditional density of the class. Fig. 4 shows the structure of the classification process in this work.

In general, The classification methods can be classified into parametric and non-parametric problems. In fact, parametric methods are based upon the assumptions of normally distributed population, and they estimate the parameters of the distributions to solve the problem. However, nonparametric methods make no assumptions about the specific distributions involved, and are therefore distribution-free [28].

TABLE II.   NORMALIZED HU MOMENTS OF TABLE I

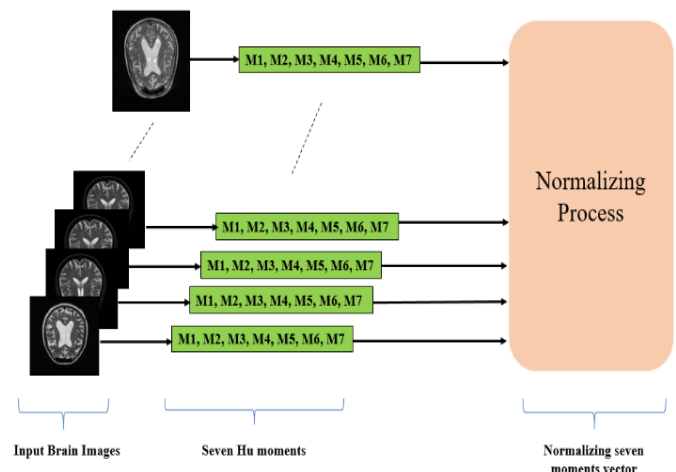| Hu moments	Image1 (CDR=0)	Image2 (CDR=1)	Image3 (CDR=2)	Image4 (CDR=3)
M1	0.4377	0.5107	0.4879	0.5564
M2	0.1181	0.5086	0.3772	0.7650
M3	0.5518	0.3602	0.1699	0.7328
M4	0.5416	0.2852	0.0391	0.7898
M5	0.1377	0.6978	0.1679	0.6826
M6	0.0776	0.0859	0.7246	0.6794
M7	-0.4362	-0.5101	-0.4867	-0.5592

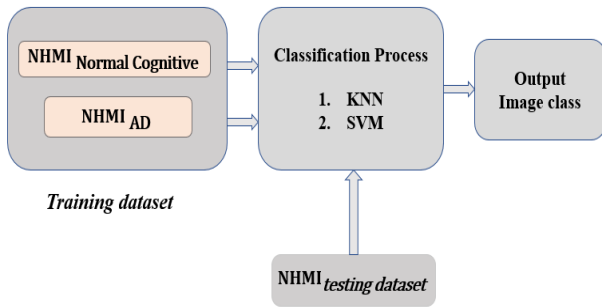Fig. 3.   Structure of feature extraction process.

Fig. 4.   Block diagram of classification process.

1) K- Nearest Neighbor KNN

The KNN classifier assists as a design of a non-parametric statistical method. When a testing data is examined, a K-NN classifier tries to find the pattern space for the k training cases which are alike in unknown cases. These k training cases consider the "K-nearest neighbors" of the unknown cases. K-NN classifier can also be suitable for the dependent variable that deals with more than two principles like high risk, medium risk, and low risk. Besides, K-NN classifier needs an identical number of good and bad sample cases for improved performance. The selection of k also infuluences the performance of the k-NN process [28].

The K-nearest neighbor algorithm depends on the knowledge of clustering components of similar nature. In other words, items of the same class should be nearer in distance [29]. The execution process of the K-nearest neighbor algorithm is as follows:

Let T be a training dataset, and S a test dataset. Individually, every sample x_a is a tuple (x_{a1}, x_{a2}, ..., x_{aD}, z), where, x_{af} is the value of the f-th feature of the a-th sample. This sample belongs to a class z, represented as xza, and a specific dimensional space. For the T set, the class z is identified, while it is unidentified for S set. Basically, for each sample x_{test} held in the S set, the k-NN model searches for the k nearest samples in the T set. Mathematically, it calculates the distances between x_{test} and all the samples of T set. Normally, the Euclidean distance is used for this task. According to this calculated distance, the k closest samples ($neigh_1$, $neigh_2$, ..., $neigh_k$) are found by placing the training samples in ascending direction. Based on the k closest neighbors, a majority vote is managed to compute which class is prime among the neighbors. The value of k could possibly affect the performance and the noise of this method [30]. So, the KNN algorithm can be summarized as two main procedures [29]:

a) First, the number of closest points of test sample x against training data T is determined using a Euclidean distance equation. If there are two points in j dimensional space, x = [x_1, x_2, ..., x_j] and y = [y_1, y_2, ..., y_j], the Euclidean distance between them can be denoted by (15) [28]:

$$d(x, y) = \sqrt{\sum_{i=1}^{k} (y_i - x_i)^2} \qquad (5)$$

b) When a test sample x has more representatives than a specific class of data, which means the number of K-nearest

points accounting for the majority, it is judged that x is of that specific class [29].

2) Linear SVM

Support Vector Machines (SMVs) were used as a classification method, using LIBSVM toolbox under MATLAB as simulation software [31]. Firstly, SVMs are expressed for binary classification. The SVM technique is a familiar model which has shown to perform perfectly in various applications by similar or improved performance than many other models. SVM has an additional benefit over other approaches. It is computationally less sensitive to the dimensionality of the application, which permits dealing with complex applications of a large number of variables [32].

SVM is a supervised learning technique. It is a binary classifier which returns a class label. SVM splits binary labels of the training data by the following hyperplane:

$$g(x) = w^{T} x + w_0 \qquad (6)$$

Where, w is known as the weight vector and w_0 as the threshold. Fig. 5 illustrates the hyperplane of a linear SVM. This hyperplane is extremely distant from the two classes [33]. The thematic of a binary classifier is to build a function *f*:

$\mathbb{R}^n \rightarrow \{\pm 1\}$ using training data that is, *n*-dimensional patterns x_i and class labels y_i:

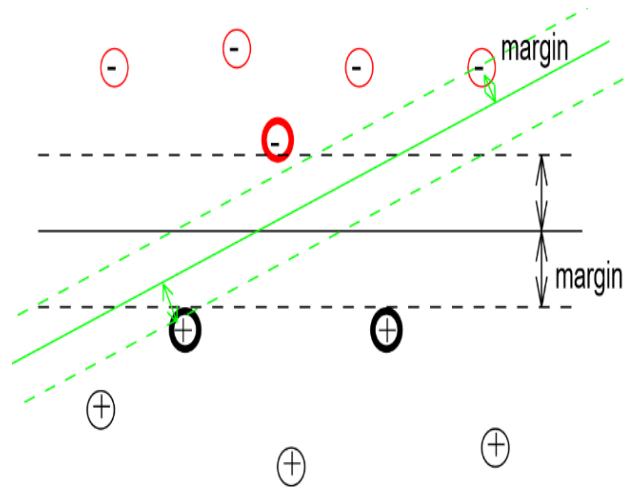$$(X_1, Y_1), (X_2, Y_2) ..... (X_n, Y_n Z) \in (R \rightarrow \{\pm 1\}) \qquad (7)$$

Fig. 5.   The hyperplane of a linear SVM.

So as to *f* will properly categorize the new samples (**x**, y) [29]. This linear separating is found with a maximum-margin in a richer feature space made by kernel function k (x, z). There are many general kernel functions consisting of polynomial, RBF, sigmoid, etc. The typical formula of SVM classifier is defined as follows [34]:

$$f(x) = \sum_{i=1}^{n} \alpha_i \, y_i \, k(x, v_i) + b \qquad (8)$$

Where, $\{v_i\}_{i=1}^{n}$ are denoted the support vectors which are a minor set of training data close to the splitting hyperplane.

V. KNN vs SVM

Several classifiers have been established by many researchers, which are used in systems that include object recognition. Practically, both K-Nearest-Neighbor (KNN) and Support-Vector-Machine (SVM) classifiers are well known and commonly used.

In KNN, the object is classified based on the labels of its k nearest neighbors by popular vote. When k=1, the object is easily classified as the class of the object closest to it. If there are just two classes, k should be an odd number. A core benefit of the KNN algorithm is its strong performance with multi-modal classes, since the base decision is built on a minor neighborhood of the same objects. So, the system can still result with good accuracy if the goal class is multi-modal. However, a main weakness of the KNN algorithm is that it uses all the features equally in calculating for similarities. This could result in classification errors, particularly when there are just few subsets of features that are valuable for classification.

KNN has some suitable properties. It is mechanically non-linear; it can recognize linear or non-linear distributed data; and it works very well with a lot of data points. On the other hand, KNN has some negatives. It needs to be carefully tuned; the selection of K and the metric (distance) to be used are crucial. Besides, KNN may be slower to use when the value of K is to be reserved high, or the total number of points is high.

A main positive of SVM classification is that SVM performs well when datasets have numerous characteristics, even if there are just a few cases that exist for the training process. SVM performs in a different way and it is a good and fast solution for many applications. But, some disadvantages of SVM classification include limits in speed and size throughout both training and testing processes of the system and the collection of the kernel function parameters. Eventually, if the application has a lot of points in a low dimensional space, then KNN is perhaps an excellent choice. If the application has a few points in a high dimensional space, then a linear SVM is possibly better.

VI. EXPERIMENTAL RESULTS

The experiments have been evaluated using two different classifiers: KNN and SVM. Alzheimer database images in this work have been classified into two classes: normal cognitive (MCI) and brain suffered from AD. The proposed approach is trained with different numbers of training databases for each classifier, and their performance in discriminating the healthy and AD brain images are investigated.

A. Training

In this stage, the system is trained using the NHMI algorithm. The feature extraction process is performed for two distinct categories. Using NHMI, the moments' values will be separated into distinct, unique sets to guarantee the performance of the classifiers during the classification process. At the end of this stage, the most important features were constringed as 1D vector, which contains normalized moments for each training set (healthy and AD).
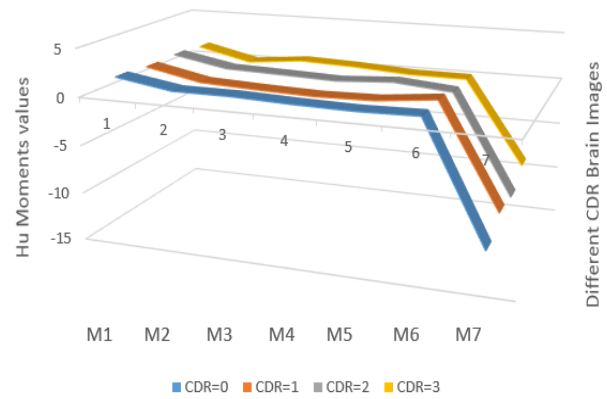

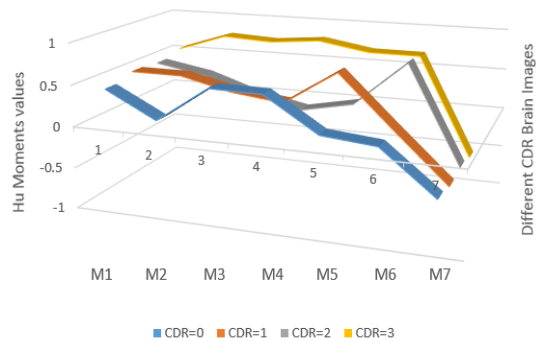Fig. 6. Hu moment invariants for four different brain images.

Fig. 7. Normalized Hu moment invariants of Fig. 5.

By way of example, Table 2 illustrates the normalized moments of four different brain images. It shows the effective power of NHMI weights. These weights represent the dominant distinct features of each image. Fig. 6 displays an example of the salient HMI features, there is a clear convergence of their moment's values which make it kind of confusing for the classifiers to do the classification tasks and put each test dataset in its right class. Therefore, the idea of normalizing these moments has been proposed to ensure the moments of every class are separated differently from each other. This step highly improves the classification system performance and the accuracy reaches 100% for the SVM classifier. Fig. 7 shows the same moments after normalization and how they look spaced out between each other, which was the key point in this work.

B. Testing and Results

With a view to accomplish best-expected accuracy, we tested the system using two sets of brain images; MCI and AD with two different classifiers, KNN and SVM. The classification accuracy is also estimated for each classifier, too. Indeed, NHMI model demonstrates an improved classification performance, as well as training, once the moments are normalized. Fig. 8 shows the training datasets distribution for SVM classifier and how the hyperplane separates the two classes non-linearly. However, the normalizing process has solved this problem and enables the classifier to recognize each class perfectly. For evaluation purposes, the testing results are approved using ADNI datasets. It is proven that the designed NHMI shows

promising results. As shown in Tables 3 and 4, confusion matrices include the classification accuracy for each used classifier. The usefulness of the normalized moments is demonstrated in Table 5, where the confusion matrices and the classification accuracy are recorded before normalizing the Hu moments, even though we use 100 subjects as a training dataset. To be fair, we run the same number of training and testing datasets for the two different classifiers. Also, the running time that is required for both classifiers to do the classification process has been computed. Table 6 displays the running time values. As it is clear in this work, KNN is faster than the SVM classifier, since we have just two classes. However, SVM performs better than KNN in the classification performance. Table 3 illustrates how the NHMI performance improved as the number of training datasets increased for KNN classifier. While for SVM, it got its perfect performance from the beginning when the training dataset is 50, so, SVM does not need any increase with the training datasets. However, each classifier has its strength in classifying the testing datasets in our model.

C. Sensitivity and Specificity

In addition to the accuracy rate of the classification approach, there are other statistical measures for a binary classification test named as sensitivity and specificity. They are widely used to describe a diagnostic test. In any medical study, each subject may have or may not have the disease. The test result can be either positive (having the disease), or negative (does not have the disease). Nevertheless, there is still a possibility that the test outcome does not match the actual case of the patient. Sensitivity calculates the ratio of actual positives which are correctly diagnosed (the percentage of the patients who are recognized to have the disease). While Specificity computes the ratio of negatives which are correctly diagnosed (the percentage of healthy people who are recognized as not having the disease). They can be expressed as follows:

$$Sensitivity = TP / (TP + FN),$$
$$Specitivity = TN / (TN + FP) \tag{9}$$

where *TP* is the number of true positives, which means number of AD patients who were correctly classified, *TN* is the number of true negatives which is the number of normal cognitive correctly classified; *FN* is the number of false negatives, the number of AD patients classified as normal cognitive, and *FP* is the number of false positives which is the number of normal cognitive people classified as AD patient. These probabilities expose the skill to distinguish MCI/AD patterns as illustrates in Table 7.
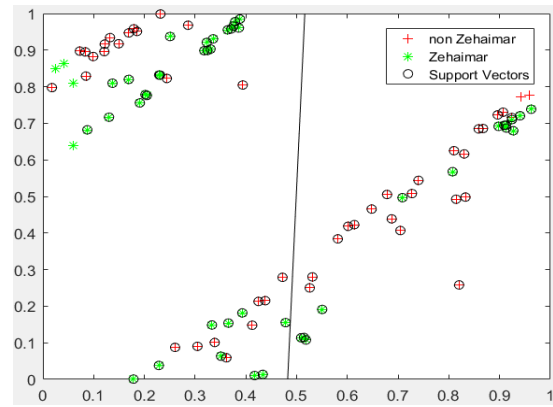

Fig. 8.   Separating Hyperplane for the input training data space using SVM classifier.

TABLE III.   CONFUSION MATRICES OF CLASSIFICATION ACCURACY FOR KNN CLASSIFIER

Brain images	Training dataset =50			Training dataset=75		
	MCI	AD	Accuracy	MCI	AD	Accuracy
MCI	16	0	100%	16	0	100%
AD	4	8	66%	2	10	83%
Average Accuracy	83%			91.5%		
MCI	Training dataset =100					
	MCI	AD	Accuracy			
	16	0	100%			
AD	2	10	83%			
Average Accuracy	91.5%					

TABLE IV.   CONFUSION MATRICES OF CLASSIFICATION ACCURACY FOR SVM CLASSIFIER

Brain images	Training dataset =50			Training dataset=75		
	MCI	AD	Accuracy	MCI	AD	Accuracy
MCI	16	0	100%	16	0	100%
AD	0	12	100%	0	12	100%
Average Accuracy	100%			100%		
MCI	Training dataset =100					
	MCI	AD	Accuracy			
	16	0	100%			
AD	0	12	100%			
Average Accuracy	100%					

TABLE V.   CONFUSION MATRICES OF CLASSIFICATION ACCURACY BEFORE NORMALIZING THE MOMENTS WITH TRAINING DATASETS=100

Brain images	KNN			SVM		
	MCI	AD	Accuracy	MCI	AD	Accuracy
MCI	11	5	68.7%	12	4	75%
AD	3	9	75%	9	3	75%
Average Accuracy	71.85%			75%		

TABLE VI.   RUNNING TIME FOR CLASSIFICATION PROCESS WITH DIFFERENT TRAINING DATASETS FOR BOTH KNN AND SVM CLASSIFIERS (MSEC.)

Training datasets	KNN	SVM
50	2.1	4.9
75	2.1	13.7
100	2.5	28.9

TABLE VII.   AVERAGE ACCURACY, SENSITIVITY, SPECIFICITY, FOR KNN AND SVM WHEN TRAINING DATASETS=100

Classifier	Sensitivity	Specificity	Average Acc.
KNN	0.833	1	91.5%
SVM	1	1	100%

D. Results Comparison

The work of the NHMI model in this paper is compared with other state-of-the-art techniques which used the same ADNI database. Table 8 clarifies the classification accuracy results of the ADNI database literature works used different algorithms, in comparison with ours. Overall, the proposed NHMI proves a considerable enhancement in performance compared with other state-of-the-art methods.

VII. DISCUSSION

As has been noted, normalizing the Hu moments significantly affects our proposed system performance and gives an outstanding result, especially for SVM classifier. SVM classifier has reached its best performance (accuracy of 100%) once we normalize the Hu moments no matter how much we increase the training datasets. From the other point of view, the KNN classifier affected with increasing the training datasets which enhances its performance, but no longer than 75 training datasets. It is quite obvious that KNN classifier has a simpler model structure, which makes it the faster classifier for low-level classification applications of the two classes, having low dimensional features like our application. On the other hand, SVM classifier has the better performance, but it is a little bit more complex and slower than KNN. As a future work, we are looking headlong to classify Alzheimer disease into four different classes, depending on measures of disease severity (normal cognitive CDR=0, simple AD (CDR=1), moderate AD (CDR=2), severe AD (CDR=3)). Besides, investigating new classification schemes to recognize and classify other diseases, ECG signals classification, or classification of various medical images for healthy and non-healthy people.

TABLE VIII.   COMPARISON CLASSIFICATION ACCURACY OF OUR NHMI WITH THE STATE-OF-THE-ART EXISTING ALGORITHMS FOR THE SAME ADNI DATASETS

ADNI literature	Classification Methods	Acc (%)
Carlos Cabral et al. [14]	Favourite Class Ensembles	66.78%
L. Herrera et al. [25]	DWT and SVM classifier	83.63%
J. Iglesias et al. [35]	Self-Smoothing Operator	97.5%
D. Zhang et al. [36]	Multimodal classification based on MRI, PET, and CSF	93.2%
M. López et al [37]	Principle Component Analysis with neural network	96.7%
This work	**NHMI with KNN classifier** **NHMI with SVM classifier**	**91.5%** **100%**

VIII. CONCLUSION

In this paper, we investigate and prove the usefulness of a NHMI technique for Alzheimer disease classification. The NHMI system serves medical requests in a specific area for diagnosing medical images using good classification techniques with fast processes. It offers greater advantages, perfect classification accuracy (100% for SVM classifier), and fast computational processes.

Hu moments Invariant algorithm has been used in this approach. The key point in this work is to normalize these moments and make them diverge from each other, which results in perfect classification performance. This step has made the proposed system very efficient, even though it is built from common classifiers. Two different classifiers have been used in this study, KNN and SVM. The experiments' results are obtained during short running time and ideal classification accuracy. To guarantee the preferred results, different numbers of the training datasets are used (50, 75, and 100), and 28 datasets are used for testing purposes later. Best results, in terms of best accuracy and low running time, are obtained with moderate numbers of training datasets.

REFERENCES

[1] S. Sarraf, G. Tofighi, and D. Neuroimaging, "DeepAD : Alzheimer's Disease Classification via Deep Convolutional Neural Networks using MRI and fMRI," Aug 2016.

[2] V. Sivakumar, V. Murugesh, and S. Arabia, "Australian Journal of Basic and Applied Sciences Alzheimer ' s Disease Detection in Brain Magnetic Resonance Images Using SOM," vol. 9, no. January, pp. 315–322, 2015.

[3] R. Mahmood and B. Ghimire, "Automatic Detection and Classification of Alzheimer's Disease From MRI Scans Using Principal Component Analysis and Artificial Neural Networks," pp. 133–137.

[4] X. Zhou, S. Wang, W. Xu, G. Ji, P. Phillips, and W. Virginia, "Detection of Pathological Brain in MRI Scanning Based on Wavelet-Entropy and Naive Bayes Classifier," pp. 201–209, 2015.

[5] T. A. Intelligence, "Pathological brain detection in MRI scanning via Hu moment invariants and machine learning," no. January, 2016.

[6] M. K. Hu. Visual pattern recognition by moment invariants. IRE Trans. Inform. Theory, IT(8), 1962.

[7] F. Al-azzo and A. M. Taqi, "3D Human Action Recognition using Hu Moment Invariants and Euclidean Distance Classifier," vol. 8, no. 4, pp. 1–9, 2017.

[8] N. Belmokhtar and N. Benamrane, "Article: Classification of Alzheimer's Disease from 3D Structural MRI Data," International Journal of Computer Applications, vol. 47, no. 3, pp. 40–44, June 2012, published by Foundation of Computer Science, New York, USA.

[9] I. Illán, J. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, F. Segovia, R. Chaves, M. Gómez-Rio, and C. Puntonet, "18F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis," Information Sciences, vol. 181, no. 4, pp. 903–916, 2011.

[10] P. Padilla, M. López, J. Górriz, J. Ramírez, D. Salas-Gonzalez, and I. Álvarez, "NMF-SVM Based CAD Tool Applied to Functional Brain Images for the Diagnosis of Alzheimer's Disease," Medical Imaging, IEEE Transactions on, vol. 31, no. 2, pp. 207–216, 2012.

[11] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," Neuroimage, vol. 55, no. 3, pp. 856–867, 2011.

[12] M. Lopez, J. Ramirez, J. Gorriz, D. Salas-Gonzalez, I. Alvarez, F. Segovia, and R. Chaves, "Multivariate approaches for Alzheimer's disease diagnosis using Bayesian classifiers," in Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE, 24 2009-nov. 1 2009, pp. 3190 –3193.

[13] J. Young, M. Modat, M. Cardoso, J. Ashburner, and S. Ourselin, "Classification of Alzheimer's disease patients and controls with Gaussian processes," in 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI), May 2012, pp. 1523 –1526.

[14] C. Cabral, M. Silveira, and N. Initiative, "Classification of Alzheimer ' s Disease from FDG-PET images using Favourite Class Ensembles," pp. 2477–2480, 2013.

[15] M. Liu, D. Zhang, and D. Shen, "Ensemble sparse classification of Alzheimer's disease," NeuroImage, vol. 60, no. 2, pp. 1106 – 1116, 2012.

[16] R. Polikar, C. Tilley, B. Hillis, and C. Clark, "Multimodal EEG, MRI and PET data fusion for Alzheimer's disease diagnosis," in Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, 31 2010-sept. 4 2010, pp. 6058 –6061.

[17] Rowayda, A. S. (2013). Regional atrophy analysis of MRI for early detection of alzheimer's disease. International Journal of Signal Processing, Image Processing and Pattern Recognisation, 6(1).

[18] Iraky, K., Aliaa, Y., & Howida, Y. (2012). MRI brain image segmentation based on wavelet and FCM algorithm iraky khalifa. International Journal of Computer Applications, 47(16).

[19] E.Bron,M.Smits,W.vanderFlieretal.,"Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CADDementia challenge," NeuroImage, 2015.

[20] S. Kl¨oppel, C. Stonnington, C. Chu, B. Draganski, R. Scahill, J. Rohrer, N. Fox, C. Jack, J. Ashburner, and R. Frackowiak, "Automatic classification of MR scans inAlzheimer'sdisease,"Brain,vol.131,no.3,pp.681– 689, 2008.

[21] Y. Fan, D. Shen, R. Gur, R. Gur, and C. Davatzikos, "COMPARE: classification of morphological patterns using adaptive regional elements," IEEE Trans. Med. Imag., vol. 26, no. 1, pp. 93–105, 2007.

[22] Khotanzad, A., and Y.H. Hong, 1990. Invariant image recognition by zernike moments. IEEE Trans. Pattern Analysis and Machine Intelligence, 12: 489-497.

[23] M. Rizon *et al.*, "Object Detection using Geometric Invariant Moment," *Am. J. Appl. Sci.*, vol. 3, no. 6, pp. 1876–1878, 2006.

[24] A. Khotanzad and J. Lu, "Classification of invariant image representation," IEEE Transaction on Acoustics, Speech, and Signal Processing, vol. 38, 1990.

[25] L. J. Herrera, I. Rojas, H. Pomares, A. Guillén, O. Valenzuela, and O. Baños, "Classification of MRI images for Alzheimer ' s disease detection," 2013.

[26] Y. Zilu, L. Jingwen and Z. Youwei, "Facial expression recognition based on two dimensional feature extraction Signal Processing", ICSP 9th International Conference on IEEE, (2008), pp. 1440-1444.

[27] V. Sugumaran, V. Muralidharan and K. I. Ramachandran, "Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing", Mechanical Systems and Signal Processing, vol. 21, no. 2, (2007), pp. 930-942.

[28] A. K. Abdelmoula, "Bank credit risk analysis with k-nearest- neighbor classifier : Case of Tunisian banks," vol. 14, no. 1, pp. 79–106, 2015.

[29] C. Chen, W. Huang, T. Tan, C. Chang, and Y. Chang, "Using K-Nearest Neighbor Classification to Diagnose Abnormal Lung Sounds," pp. 13132–13158, 2015.

[30] I. Triguero, "A MapReduce-based k-Nearest Neighbor Approach for Big Data Classification," 2015.

[31] Chan, C., & Lin, C. (2011). LIBSVM: a library for support vector machines. ACM Transactions on Intelligen Systems and Technology, 2 (3), 2:27:1-27:27.

[32] L. J. Herrera, I. Rojas, H. Pomares, A. Guillén, O. Valenzuela, and O. Baños, "Classification of MRI images for Alzheimer ' s disease detection," 2013.

[33] M. Lo, "Neurocomputing Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer ' s disease," vol. 74, pp. 1260–1271, 2011.

[34] H. Cao et al., "Approximate RBF Kernel SVM and Its Applications in Pedestrian Classification To cite this version : Approximate RBF Kernel SVM and Its," 2008.

[35] J. E. Iglesias, J. Jiang, C. Liu, and Z. Tu, "Classification of Alzheimer ' s Disease Using a Self-Smoothing Operator," pp. 58–65, 2011.

[36] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "NeuroImage Multimodal classification of Alzheimer's disease and mild cognitive impairment," vol. 55, pp. 856–867, 2011.

[37] M. Lo, "Neurocomputing Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer ' s disease," vol. 74, pp. 1260–1271, 2011.