Mathematical Model and Artificial Intelligence for Diagnosis of Alzheimer's Disease

Afsaneh Davodabadi
Islamic Azad University of Tehran: Islamic Azad University Central Tehran Branch

Behrooz Daneshian (✉ Be.daneshian@iauctb.ac.ir)
Islamic Azad University Central Tehran Branch

Saber Saati
Islamic Azad University Tehran North Branch

Shabnam Razavyan
Islamic Azad University South Tehran Branch

Research Article

Keywords: Mental state, diagnosis, classification, deep learning, SVM, DCNN, Mathematics

Posted Date: January 9th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2365650/v1

License: ☕️ 🔢 This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Degeneration of the neurological system linked to cognitive deficits, daily living exercise clutters, and behavioral disturbing impacts may define Alzheimer's disease. Ad research conducted later in life focuses on describing ways for early detection of dementia, a kind of mental disorder. To tailor our care to each patient, we utilized visual cues to determine how they were feeling. We did this by outlining two approaches to diagnosing a person's mental health. Support vector machine is the first technique (SVM). Image characteristics are extracted using a fractal model for classification in this method. With this technique, the histogram of a picture is modeled after a Gaussian distribution. Classification was performed with several SVM kernels, and the outcomes were compared. Step two proposes using a deep convolutional neural network (DCNN) architecture to identify Alzheimer's-related mental disorders. According to the findings, the SVM approach accurately recognized over 93% of the photos tested. The DCNN approach was one hundred percent accurate during model training, whereas the SVM approach achieved just 93 percent accuracy. In contrast to SVM's accuracy of 89.3%, the DCNN model test's findings were accurate 98.8% of the time. Based on the findings reported here, the proposed DCNN architecture may be used for diagnostic purposes involving the patient's mental state.

1. Introduction

People all throughout the globe face challenges related to the availability, accessibility, cost, and quality of health care. Many factors, including socioeconomic status and geographic location, contribute to the wide disparities in people's access to and ability to afford health care services[1]. Many individuals are now let down by the health care system. Providing sufficient health care is a problem in many nations as well[2]. Some examples of these difficulties include [3] the time it takes for doctors to see patients, [4] the difficulty in sharing information across different medical systems, [5] and the high cost of [6] medical technology and infrastructure. The increasing rates of chronic illnesses in both advanced and emerging countries further highlight the need of using cutting-edge technological solutions. The global population is also sometimes confronted with a major threat from infectious illnesses. One difficulty is the shortage of medical professionals[7], [8]. Countless individuals have suffered because of anything like the proliferation of sickness. One of the greatest ways to get to patients quickly and effectively in such situations is via patient support systems. Fortunately, these difficulties may be overcome with the help of clever portable technologies[9]. Health care may be enhanced by the use of mobile apps, sensors, drugs, and remote patient monitoring products[10]. By streamlining the delivery of care services and goods and improving lines of communication between patients and medical professionals, these innovations have the potential to significantly cut down on health care expenditures. Applications powered by AI have the potential to improve patient and medical professional education, chronic disease management, and the tracking of key health metrics[11]. One of the emerging and crucial problems in medicine is how to tell someone's health only by looking at their face. Previous research has shown that such facial expressions may be utilized as a significant indicator in the diagnosis of mental health issues such schizophrenia, depression, autism, and bipolar disorder[12], among others[13]-[16]. An unhealthy relationship with
unpleasant emotions, an inability to express feelings, or emotional instability have all been linked to the aforementioned disorders. Human analysis in picture analysis is an intriguing problem in machine vision. Human computer interaction, emotion analysis, interactive video, display and retrieval from picture and video databases, and computer-generated face animation are just some of the many applications of face recognition technology. The human face may be utilized to convey non-verbal and conceptual information to a computer, including a person's emotional and spiritual state[17]. Light levels, head orientation, and even subtle expression shifts may all distort a face picture. Due to these characteristics, face identification remains a difficult problem[18]. The term "morphological operation" is taken from the study of form and structure in living organisms. In order to convey traits and describe the form of regions, such as border regions, frames, and convex shells, mathematical morphology is employed[19]. When applied to image processing issues, the unification of mathematical morphology-based language, set theory, and morphology provides a strong and effective solution[20]. Morphology with a mathematical foundation uses sets to express the features of an image. A fractal model is introduced that uses characteristics including texture, light intensity, and edge to do segmentation and feature extraction. This technique is derived by mining the data for the relevant insights. That's why [[21]] we choose just intrinsic characteristics for classifying data.

8% of Westerners over the age of 85 get Alzheimer's disease (AD), making it the most frequent and most disabling illness in the opinion of every neurologist [1]. There are monoclonal and familial variants of this illness [2]. The single clamp style is the most practical option. Late-onset Alzheimer's disease (typically beyond age 65) is characterized by gradual atrophy of the cerebral cortex, particularly in the frontotemporal and parietal lobes, and enlargement of the lateral and third ventricles. The cerebral cortex widens, and the grooves in it widen as well [3–5]. Some additional signs of Alzheimer's disease include a decrease in brain cell count as well as shrinkage and distortion of the surviving neurons. Extracellular aging plaques, intracellular neurofibrillary tangles, and vascular amyloid deposits are pathological hallmarks of this illness [6, 7]. The Electroencephalogram (EEG) was first proposed in the late 30s of the last century [1] as a means of recording and subsequently giving information about human brain activity using electrical stimulation of the scalp. Recently, it has been used as a tool for detecting Ad. However, activities containing polymorphic moderate waves in the theta or delta range, frontal irregular musical delta action (FIRDA), and other EEG findings in such people are not unique, and any measurement of this movement was not feasible a few decades ago [8]. This was about to be remedied with new approaches, and now the hallmarks of EEG deviations from the norm in Ad patients are as follows: a shift of the control range to frequencies with lower rate, a diminish incoherence of quick rhythms, and EEG-complexity changes, that are found as of now within organize in advance, in a wide recurrence extend [9–12]. Cortical nerve cell pathology (damage to the long threadlike part of a nerve cell along which impulses are conducted from the cell body to other cells), axon pathology (damage to the long threadlike part of a nerve cell along which impulses are conducted from the cell body to other cells), and a lack of cortical nerve cells in which acetylcholine acts as a neurotransmitter are all suspected causes of these abnormalities.
In this work, we explore how a smart environment may automatically make use of pictures and emotions to improve healthcare. To tailor our care to each patient, we utilized visual cues to determine how they were feeling. To do this, we will be employing the Support vector machine for classification after extracting the fractal characteristics from the photos. The patients' mental conditions have also been determined using the robust convolutional neural network (CNN) technologies. As this post progresses, we will discuss the literature study, our methods, and our CNN's underlying architecture. In the results and discussion section, we'll go through the findings in further depth.

2. Literature Review

Alzheimer's disease is linked to a decline in the capacity of the unconscious to retrieve memories and ideas. It's a mental disorder where the patient's capacities deteriorate with time [15]. Memory loss is the hallmark symptom of dementia [16]. Most cases of memory loss happen gradually and worsen over time. At first, only current experiences and lessons are affected by memory loss, but with time, even long-term memories are affected [17]. The patient fails to retain information and promptly repeats a question that was just answered. The patient cannot find his personal items and has no idea where he put them. He has trouble making and receiving monetary transactions and is unable to keep track of his financial holdings. It becomes more difficult to identify and name familiar faces [18–21]. The navigational issue is slowly being uncovered, and if left unattended, it might be forgotten. The bedroom, kitchen, bathroom, and toilet in his own house may no longer be familiar to him [22–25]. Due to the disruption in the patient's ability to think and remember, the patient may experience feelings of sadness, rage, and violence. Hallucinations and delusions are a common symptom of dementia. The patient may falsely believe, for instance, that his wife has betrayed him or that his neighbors and nurse are plotting against him. Patients may have a negative outlook towards fatherhood. Sometimes the patient sees persons, such as departed parents or relatives who are not present [26–28].

A serious illness sometimes necessitates that the patient be assisted with even the most fundamental of daily activities. A patient with dementia may have trouble communicating and finding the correct words, resulting in dullness and isolation. In severe circumstances, the patient loses awareness of his illness and may act recklessly since he is unaware of his limitations. The patient's mobility may deteriorate with time, and he may have frequent falls due to poor balance [29].

2-1- Methods of Alzheimer's identification and medical care

Pre-disease reactions in the body may be used to infer the presence or absence of a disease. Among the biomarkers used in the Alzheimer's disease diagnostic process are imaging modalities of the brain. The reduction in biomarker fuel for this patient is shown in Fig. 2 by reduced metabolism in the brains of Alzheimer's patients in the temporomandibular area and above the skull, as detected by PET-FDG4. [30–33]

The term "biomarker" refers to any biological marker that may be measured and utilized as an indicator of the presence or absence of a condition, such as a disease [34]. The levels of beta-amyloid and proteins
called tau in the CSF fluid are one of the first indicators used to identify Alzheimer's disease [35]. This is because the key alterations in Alzheimer's patients are beta-amyloid acid deposition and tau protein accumulation in the brain. Predictive validity in cerebrospinal fluid testing is rather strong. Patients with Alzheimer's disease have consistent levels of beta-amyloid protein throughout time, and the pace of decline is inversely proportional to the rate of disease development [36]. Neuronal synaptic injury is indicated by increased levels of tau phosphorylated protein or total tau protein. Alzheimer's patients tend to have elevated end-tau levels, however this may not be indicative of the severity of the illness [37].

There has been lot of study into face recognition over the last several decades. Some of them include approaches based on detecting motor units [22], statistics techniques [23], feature extraction [24], local binary pattern operators [25], mathematics transformations including discrete wave transformations [26]–[28], Gabor filters [29], etc. and hybrid methods [30]–[33]. An ICU patient monitoring study by Davoudi et al. [34] was recently published in nature. They employed wearable light and sound sensors, and a camera to gather patient's face pictures in the surroundings. They gather patients, facial expressions, postures, and other out-body reactions. Results reveal that there is a considerable difference among delirious and non-delirious individuals' facial expressions. Patients' facial expressions were also modeled using a deep learning method. The findings further suggest that patients are affected by their surroundings.

Narayanan et al. [35] utilized online Face2Gene software to detect dysmorphic disorders. They employed 51 patient facials were examined using facial analysis techniques. Computer-based technologies for the diagnosis of genetic disorders offer promising results. Using Face2Gene, the researcher also looked at congenital dysmorphic diseases in previous work published in nature. They examine two groups of individuals from Japan with congenital dysmorphic disorders. The study found that face analysis is accurate in 82.2% of diagnoses [36]. Denadai et al. [37] explored orthognathic surgery utilizing a 3D face imaging panel. To determine causal causes, they examined psychological indicators. The findings validated the use of facial expression analysis as a therapeutic tool in orthognathic surgery. Orthognathic surgery has been studied for its effects on both face expression and mental health ([38]). Computer-aided facial expression categorization was given by Wu et al. [39]. They employed a convolutional neural network to diagnose normal and abnormal faces. Ashraf et al. [40] proposes an automated face recognition system for the diagnosis of pain for patients in ICU. The employed photos without their occlusions, however, impose masks on patients' faces. Images of faces with open eyes serve as input. Patients' ocular characteristics were employed in a computer-based categorization system. The approaches outlined worked well. Using video footage of an interview, Bishay et al. [41] examined the patient's facial expressions to see if they were indicative of schizophrenia. They demonstrated the use of a neural network structure for symptom prognosis. Their provided network can appropriately identify the patient's symptoms. Medina et al. explored the biopsychosocial aspects of the pain of patients. Patients with alexithymia were shown a video with emotional undertones. Their research demonstrates that the nasal temperature drops when a patient expresses anger or discomfort [42].

The use of computer vision for the purpose of recognizing facial expressions has been the subject of a great deal of study. Lopes et al. [43] investigated the use of facial expression in AI. They offered the CNN approach for accomplishing their aims. Their key problem is boosting the detection accuracy. Their
suggested framework may be used to categorize how people express themselves emotionally. Similarly, some scholars provide models for expressing emotions, such as the six-edged hexagonal model for displaying facial expressions. With the display of the model, writers extract face features[44]. Ding et al. [45] introduced a double local binary pattern for face detection. They claim that an automated model has two properties: 1) peak expression frame identification and 2) expression feature extraction. The suggested solution properly reduces detection time while maintaining a low dimension. Moreover, to tackle the lighting challenges in LBP, logarithm-Laplace techniques are employed to boost the face robustness of features. Taylor expansion was eventually used to extract face emotion elements. The Taylor feature pattern (TFP) was also provided to identify a significant face feature. Their conclusion shows that the given TFP approach is superior than some of the earlier LBP-based feature extraction methods for face expression[45]. A local neighborhood difference binary pattern was suggested by Saurav et al.[46] (LNDBP). LNDBP's primary benefit is that it generates binary patterns that reflect the interdependence of all nearby points. So they have introduced LNDBP techniques to take use of both LNDP and LBP. The high dimensionality of the derived LNDBP characteristics necessitated the use of a dimension reduction technique. The kernel extreme learning machine is then used to model the simplified characteristics (K-ELM). To validate the performance of their approaches, they employed two separate databases. Using a regularized sparse representation model, Liu et al. [47] proposed a semi-supervised deep learning model to extract face characteristics. Combining deep learning with sparse representations features leverages the strengths of two formidable methods of training and recognition. The results showed that the subspace features of the characteristics retrieved by deep learning are consistent with the subspace hypothesis of the sparse technique of face recognition. Their difficulty is in improving the precision of their detecting methods. The summary of some state-of-art is illustrated in Table 1.

Table. 1 Literature review of facial expression analysis
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Method</th>
<th>Aim</th>
<th>Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fei et al.[48]</td>
<td>2020</td>
<td>Convolutional neural networks with deep layers</td>
<td>Assessment of Feelings Regarding Mental Health Services</td>
<td>Accuracy and effectiveness at a high level</td>
</tr>
<tr>
<td>Akturk et al. [49]</td>
<td>2020</td>
<td>Oscillations in the Brain Measured by EEG</td>
<td>Rhythmic brain activity correlated with events measured by EEG</td>
<td>reduced phase locking and delta responses in the occipital cortex</td>
</tr>
<tr>
<td>Zhao and Lu[51]</td>
<td>2020</td>
<td>Neural Network with Deep Convolution Layers</td>
<td>System for Efficient Diagnosis of Autism</td>
<td>able to make early diagnoses of autism</td>
</tr>
<tr>
<td>Denadai et al.[37]</td>
<td>2019</td>
<td>3D Face Mapping using FACE-Q</td>
<td>evaluate corrective jaw surgery</td>
<td>The FACE-relationship Q's to other evaluation methods</td>
</tr>
<tr>
<td>Chen et al. [52]</td>
<td>2019</td>
<td>Deep convolutional neural network</td>
<td>Specifically, the Harr feature is used by the face detector.</td>
<td>Extremely precise</td>
</tr>
<tr>
<td>Matsuoka et al.[53]</td>
<td>2019</td>
<td>Techniques that evolve</td>
<td>Facial prosthesis made with 3D printing</td>
<td>exactness that is quite high</td>
</tr>
<tr>
<td>Rabinson and Baiungo[54]</td>
<td>2018</td>
<td>Research in a lab setting</td>
<td>Rehabilitation of patients with facial palsy using many approaches</td>
<td>relaxation of face muscles and enhanced motor control</td>
</tr>
<tr>
<td>Ilyas et al. [55]</td>
<td>2018</td>
<td>Technology based on deep learning</td>
<td>Emotion Expression for Patients With Traumatic Brain Injury</td>
<td>Superior facial features lead to enhanced effectiveness.</td>
</tr>
<tr>
<td>Rabhi et al. [56]</td>
<td>2018</td>
<td>Interaction between machines and humans, and artificial neural networks</td>
<td>Wheelchair that can be operated by facial expressions for the disabled</td>
<td>The precision of classification</td>
</tr>
<tr>
<td>Rodriguez et al.[57]</td>
<td>2017</td>
<td>Memory Circuits That Can Store Information Temporarily</td>
<td>Understanding facial expressions of suffering</td>
<td>Boost your results beneath the curve.</td>
</tr>
<tr>
<td>Yolcu et al. [59]</td>
<td>2017</td>
<td>Deep learning</td>
<td>Analysis of facial expressions for the diagnosis of neurological illnesses</td>
<td>Enhance precision</td>
</tr>
</tbody>
</table>
3. Methods And Materials

3.1. Conceptualization

Biometric systems that employ facial recognition to identify or verify a person's identity do so by comparing the individual's unique physiological traits to a database of known face expressions. Several methods have been proposed to better recognize faces. Our primary objective is to provide a means of diagnosing a patient's mental health simply by analyzing their facial expressions. Fig. 1 depicts the theoretical framework of the approaches suggested. The first thing to do when making a conceptual diagram is to use the fractal technique to draw out characteristics from the images. After that, patient photos with accompanying facial expressions were characterized using one of two methods. One approach utilizes a support vector machine (SVM) to categorize facial emotions by extracting characteristics in a fractal fashion. The approaches have also been compared using various SVM kernels. The second method is a computational framework called a convolutional neural network (CNN) that can distinguish between different mental states only by looking at a person's emotions. This work employs the fractal model for feature extraction. Using an artificial neural network (MLP), Hassantabar et al. (2020) developed a fractal technique for analyzing COVID-19 pictures for the purpose of patient infection diagnosis [60]. The primary distinction is that just one SVM model is used for facial expression classification. In addition, many SVM kernels were tested to see which one performed best when used for classification. After that, patient photos with accompanying facial expressions were characterized using one of two methods. One approach utilizes a support vector machine (SVM) to categorize facial emotions by extracting characteristics in a fractal fashion. The approaches have also been compared using various SVM kernels. The second method is a computational framework called a convolutional neural network (CNN) that can distinguish between different mental states only by looking at a person's emotions. This work employs the fractal model for feature extraction. Using an artificial neural network (MLP), Hassantabar et al. (2020) developed a fractal technique for analyzing COVID-19 pictures for the purpose of patient infection diagnosis [60]. The primary distinction is that just one SVM model is used for facial expression classification. In addition, many SVM kernels were tested to see which one performed best when used for classification.

3.2. Fractal feature extraction

In order to extract features, it is necessary to streamline the techniques used to narrowly define a large dataset[61]. The fractal model is used to generate the characteristic in this investigation. The collection of characteristics was used to reduce the feature size and to determine the most crucial traits for making accurate category distinctions [62]. The dimensionality reduction and Eigenvalue generation in the picture were accomplished with the help of covariance analysis and the fractal technique. In the fractal algorithm, the input vectors need to have the same size and have a representation described as a 2-dimensional matrix. There is a hard requirement that high-resolution photos be grayscale. The objects are
assembled using an MN matrix, where M is the number of objects in each picture and N is the number of points in each image. And M stands for the overall size of the picture database. The average picture must be measured in order to compute an STD from any major input matrix. From there, we may calculate the covariance matrix and its eigenvalues and eigenvectors. In the fractal algorithm, M indicates the number of instructional images. And \( F_i \) will show the image's average and \( l_i \) will indicate each \( T_i \) vector image. There are initially M images, each of which contains the \( N \times N \) dimension. An N-dimensional space can exhibit each image, which is computed as Eq. (1) and Eq. (2) indicates averaging operation[63].

\[
A = N \times N \times M
\]

1

\[
F_i = \frac{1}{M} \sum_{t=1}^{m} T_t
\]

2

A huge problem in the fractal techniques determined by Eq. (3) and the covariance matrix of Eq. (4) is called the specification of the standard deviation.

\[
Variance = \frac{1}{N} \sum (x - \mu)
\]

3

\[
Cov = AA^T
\]

4

Where \( A = [Variance_1, Variance_2, \ldots, Variance_n] \) and \( Cov = N^2 \times N^2 \) and \( A = N^2 \times M \). Therefore, \( Cov \) is large value that Eigen values of \( Cov \) are evaluated by Eq. (5).

\[
U_i = AV_i
\]

5

The final step is choosing a vector with Eigen. A sequence of properties of a state-inherent function as \( N(N = 213) \) in \( \{x_1, x_2, \ldots, x_N\} \) belonging to the \( C(C = 7) \).

The class can be found on \( \{L_i| i = 1, 2, \ldots, C\} \). The fractal method aims to map data to f-dimension space that \( f < d \). At \( y_i = R^f \) the new vector feature is placed. Scattered class matrices are covariance tables measured by Eq. (6) and Eq. (7)[63].

\[
S_T = \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T
\]

6
Where the value for all the samples is $\mu$ and $\{w_i| i = 1, 2, \ldots, f\}$ is a set of the eigenvector of $f$-dimension of $S_T$ that is associated with the largest eigenvalue $f$. Samples in the new space are $y = W^T x$, which is $W_{Fractal} \in \mathbb{R}^{f \times d}$ ($170 \times d$).

3.3. Concept of Deep convolutional neural network (DCNN)

One of the most important deep learning methods is the convolutional neural network (CNN). Primary convolutional layers, Maxpooling layers, fully connected layers, and additional layers implementing different characteristics are common components of a convolutional neural network (CNN). Each system's preparation consists of two stages: forward and reverse progression [64]. The process begins with information entering the input layer, continuing through the concealed layer, and finally exiting the output layer. The initial phase of the backpropagation technique is for the input picture to feed the network. At the completion of the process, the error value is calculated. As well as changing the network weight and the cost functions diagram, this value is then sent back into the network (see Fig. 2). The hidden sub-layers that make up a CNN are of several sorts, as can be shown in the following discussion [64]:

The foundation of every convolutional network is the convolutional layer. The output of this layer is a three-dimensional neuron matrix. Convolutional neural network (CNN) transformation of input picture and central function mappings need various kernels in certain layers. There are three significant benefits to the convolution procedure:

• The number of nodes in each function diagram is reduced using the weight-sharing approach.

Since the local link between neighboring pixels is known, the target's location may be adjusted to achieve a state of equilibrium. Activation functions are a key component of neural networks, and are responsible for producing the desired results. There are many other activation functions that may be used in neural networks, but the two most important are the Tanh and the Sigmoid. Input data (- - +) are transformed by the sigmoid function into the range $[0, 1]$. Tanh data ranges from $-1$ in terms of production value. Recently, the ReLU function has been presented as an alternative activation function. ReLU is a generalized activation function $g$. The goal is to challenge the network in a way that is not linear. Using this feature, every negative pixel value is changed to zero (64).

Max pooling: The usage of Maxpooling in CNN has a number of repercussions. By using it, CNN may first specify the goal with little changes to the matrix. Second, the large scale of the picture aids CNN in identifying characteristics. CNN Maxpooling is utilized to perform a subtraction-based summation of the function during the sampling phase, allowing for access to intermediate and advanced levels of CNN. If we want to keep this knowledge, we will need to start sharing what we have. The two most typical types of pooling are the Maximum and the Average. Data preparation, preprocessing, and data improvement
are among the most underappreciated challenges. This step, however, isn't usually required. You should determine whether your work requires pre-processing before beginning any kind of data processing [64].

### 3.4. Support vector machine (SVM)

Suppose we have the set of data points \( \{(x_1, c_1), (x_2, c_2), \ldots, (x_n, c_n)\} \). And we want to classify them into two classes \( c_i = \{1, -1\} \). Each \( x_i \) is a \( p \)-dimensional vector of \( \mathbb{R} \) that is the features. For data separation between two classes, the SVM classification algorithm identifies optimal hyperparameters. The key is to choose the appropriate delimiter. The most distant divider between the two groups is the one being discussed. Its boundary points span not one but two classes, making it the most comprehensive [65]. The mathematical formulas of these two parallel hyperparameters that make up the separator boundary are shown in Eqs. (8) and (9) [65]:

\[
8 \quad w \cdot x - b = 1
\]

\[
9 \quad w \cdot x - b = -1
\]

If the training data are linearly separable, then the distance between the two parallel hyperparameters may be maximized by choosing the two border hyperparameters in such a manner that there is no data between them. Using geometric theorems, the distance between these two hyperparameters is \( 2/|w| \), So should \( |w| \) minimized. In addition, a mathematical constraint must be added to the formal specification to ensure that no data points are ever brought inside the border. For each \( i \), the following constraints are applied to ensure that no point is placed on the boundary[66]:

First class \( w \cdot x_i - b \geq 1 \) (10)

Second class \( w \cdot x_i - b \leq -1 \) (11)

The following equations can be shown as follows:

\[
12 \quad c_i (w \cdot x_i - b) \geq 1, 1 \leq i \leq n
\]

Therefore, the optimization problem is defined with minimizing \( w \), taking into account the Eq. (12) limitations.

### 3.5. Performance analysis

Accuracy (ACC), Re-call or sensitivity (R), Specificity(S), F1-score, and geometric mean sensitivity and specificity (Gmeans) are used in this analysis to determine the effects of the classification. A system's ability to discover new patients is measured by its Recall and accuracy rate. Each class's overall
performance, including classification accuracy and recall, is measured using the F1. This section defines these standards.

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (13)
\]

\[
\text{Specificity} = \frac{TP + TN}{TP + FN} \quad (14)
\]

\[
\text{Sensitivity} = \frac{TP + FN}{TP + FP + FN + TN} \quad (15)
\]

\[
\text{F1 score} = 2 \cdot \frac{PR}{PR + DR} \quad (16)
\]

\[
\text{Gmeans} = \sqrt{\text{Sensitivity} \cdot \text{Specificity}} \quad (17)
\]

The performance matrix helps to characterize the efficiency of the classification model on several categorized datasets.

### 4. Results And Discussion

#### 4.1 Data collection

There are a total of 213 photos in the Japanese Female Facial Expression (JAFFE) collection, depicting 7 different emotions (6 different expressions + 1 neutral expression). The models for these photographs are 10 different women from Japan. Sixty Japanese individuals rate each picture on six different emotional scales. Three people, Michael Lyons, Miyuki Kamachi, and Jiro Gyoba, are responsible for the creation and compilation of this database. Several scientific studies have made use of JAFFE pictures. Grayscale photos at a resolution of 256x256 are provided in uncompressed.tiff format.

The main dataset contains 213 images so in this paper for increasing accuracy and preventing possible overfitting we used data augmentation techniques with some filters, histogram, and adding noise to the image to increase dataset size. Therefore, each mental state includes 300 images. Data set volume augmented to 2100 images.

#### 4.2 Feature extraction using the fractal model

An example of a feature extraction diagram for one of the photos is shown in Figure 4. Figure 4’s blue histogram represents the frequency distribution of picture pixels. Finding the optimal normal functions for modeling the histogram is the focus of the fractal approach. From this, we may estimate the image's histogram with a sum of three red functions, each of which is represented by a green Gaussian function. Coefficients of standard green functions are equivalent to those of the resulting images in terms of their
attributes. In light of this, we have catalogued four characteristics for every picture. If we can use these four operations to create the primary picture, it will look like.

### 4.3 Implementation of classification using SVM kernels

In this subsection, facial expressions are identified using the support vector machine (SVM) technique. In this case, the JAFFE dataset was mined for information. Three hundred face photographs were assessed for each scenario, for a total of 2100 images analyzed (75% for model training and 25% for testing and verification). Each of the seven emotions (anger, disgust, fear, happiness, neutral, sadness, and surprise) were represented by one of seven possible labels (1–7), with each image having four features as independent variables.

Different kernels were also employed to training the model and the outcomes have been compared. The following are the outcomes of the classification using a polynomial kernel of degree 5. Figure 6 illustrates the confusion matrix in training data. In this picture, the diagonal diameters show the number of accurate classifications by the trained network. Table 2 provides the results of the performance assessment criteria for the suggested technique.

<table>
<thead>
<tr>
<th></th>
<th>Polynomial</th>
<th>RBF</th>
<th>Linear</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Train</td>
<td>Test</td>
<td>Train</td>
</tr>
<tr>
<td>Accuracy</td>
<td>69.20</td>
<td>68.63</td>
<td>89.83</td>
</tr>
<tr>
<td>Specificity</td>
<td>83.03</td>
<td>79.38</td>
<td>89.22</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>13.12</td>
<td>18.54</td>
<td>67.83</td>
</tr>
<tr>
<td>F1 score</td>
<td>15.63</td>
<td>9.55</td>
<td>62.22</td>
</tr>
<tr>
<td>Gmeans</td>
<td>23.85</td>
<td>26.37</td>
<td>73.63</td>
</tr>
</tbody>
</table>

Now we'll evaluate the SVM method's various kernels side by side. Table 2 summarizes the outcomes of using polynomial kernels, radial basis functions (RBFs), and linear kernels for this purpose. The training data comparison table shows that the polynomial approach with order 5 meets the most requirements. The findings show that the RBF model's SVM accuracy is 93.18 percent for the training data and 89.33 percent for the testing data. The RBF kernel has stricter performance requirements than competing techniques. It has a 96% sensitivity for diagnosing a mental state, but only an 86.67% sensitivity for a Linear kernel and an 86.03% sensitivity for a polynomial kernel.

### 4.4 Implementation of classification using DCNN
We'll use a deep convolution neural network to analyze facial expressions and label them accordingly. The approach presented for the deep convolutional network algorithm as according Figure 8 consists of five layers, which are:

Human face photos from the JAFFE collection, which comprises 2100 images in 7 Face mode (6 face modes + 1 neutral) from 10 Japanese female models. An picture of size 256x256 with 7 face modes is used to create a 4D matrix of size 2100 by 2100 in this layer.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Image Input</td>
<td>256x256x1 images with 'zero center' normalization</td>
</tr>
<tr>
<td>2</td>
<td>Convolution</td>
<td>8(3x3) convolutions with stride [1 1] and padding 'same'</td>
</tr>
<tr>
<td>3</td>
<td>Batch Normalization</td>
<td>Normalization</td>
</tr>
<tr>
<td>4</td>
<td>ReLU</td>
<td>Rectifier</td>
</tr>
<tr>
<td>5</td>
<td>Pooling</td>
<td>2x2 max pooling with stride [2 2] and padding [0 0 0 0]</td>
</tr>
<tr>
<td>6</td>
<td>Convolution</td>
<td>16(3x3) convolutions with stride [1 1] and padding 'same'</td>
</tr>
<tr>
<td>7</td>
<td>Batch Normalization</td>
<td>Normalization</td>
</tr>
<tr>
<td>8</td>
<td>ReLU</td>
<td>Rectifier</td>
</tr>
<tr>
<td>9</td>
<td>Pooling</td>
<td>2x2 max pooling with stride [2 2] and padding [0 0 0 0]</td>
</tr>
<tr>
<td>10</td>
<td>Convolution</td>
<td>32(3x3) convolutions with stride [1 1] and padding 'same'</td>
</tr>
<tr>
<td>11</td>
<td>Batch Normalization</td>
<td>Normalization</td>
</tr>
<tr>
<td>12</td>
<td>ReLU</td>
<td>Rectifier</td>
</tr>
<tr>
<td>13</td>
<td>Fully Connected</td>
<td>7 fully connected layer</td>
</tr>
<tr>
<td>14</td>
<td>SoftMax</td>
<td>SoftMax</td>
</tr>
<tr>
<td>15</td>
<td>Classification Output</td>
<td>Cross entropy</td>
</tr>
</tbody>
</table>

Convolution occurs in the second layer, and it employs 8(3x3) filters to iteratively process the input matrix. In this research, the convolution matrix is constructed by adding a border of zeros to the input matrices. The array is analyzed in one pass by the filter as well. Matrix transformation is used in the third layer of normalization. The ReLU function, often seen in convolution networks, is used in the activation function layer, the fourth layer of a convolution network. After each layer of convolution, this function clears the matrix of any negative values and passes on its positive values to the subsequent layer. Pooling occurs on the sixth layer. This layer is a sampling technique for a convolutional neural network that functions on the pixel space and is capable of maximizing within the filter range. These subsequent levels are
organized into three distinct classes. Figure 8 and Table 3 show that a completely linked layer does not appear until the thirteenth iteration. When this layer is applied on top of a matrix, the result is a vector that may be evaluated by the output layer. If you want to be technical, it creates 7 submatrixes (number of categories). The SoftMax layer, located at position 14, is the last output layer of a multi-layer classification neural network. The SoftMax function takes in values from a number of different rankings and outputs a single value between zero and one that is the total of all the ranks. To quantify the degree of inaccuracy associated with a given value of the dependent variable, the classification layer is used as the study's output layer. This layer influences when we come closer to the most accurate result. The MSE criteria, or average squared error, is utilized here. The repeat loop runs until it obtains the desired outcome. Finally, the output is identified by 7 labels with the names of the face expressions that decide the result.

The RMSE error generated by the DCNN software is shown in Figure 9. There were two epochs, each with 150 periods, and 300 iterations in all. On the test set, which comprises 30% of the whole data set, the ultimate accuracy of this study is 0.99. In addition, the final iteration has nearly little data loss.

In this study, we evaluate the computational endpoint based on the convergence slope, which must be horizontal (zero) in Figure 9's errors and losses perspective. The training data confusion matrix is shown in Figure 10. The figure's diagonal diameters represent the percentage of instances when the trained network correctly classified the data. Example: all 300 photos were accurately labeled as depicting rage. All the photos have been properly detected and trained, i.e. with a hundred percent accuracy and sensitivity, thanks to the confusion matrix of training data. Nonetheless, the test sample confusion matrix is shown in Figure 11. Diagnostic accuracy is close to perfect, with the lowest rate being for depression. Only one out of ninety randomly chosen photographs is incorrectly categorized as cheerful. Consequently, we can say with 99.8 percent certainty that the tests have been accurate.

5. Conclusion

In this work, we explore how a smart environment may automatically make use of pictures and emotions to improve healthcare. To tailor our care to each patient, we utilized visual cues to determine how they were feeling. To do this, we employed the Support vector machine for classification after extracting fractal characteristics from the photos. There have been experiments done in this area, and it has been possible to accurately identify each mental state one hundred percent of the time. Good convergence was obtained using the suggested DCNN model, and the accuracy of the test data was close to 98%. Patients' mental health may also be assessed using the support vector machine technique. The JAFFE was mined for information. Two hundred and ten facial photos were assessed for each scenario, with seventy-five percent used for model training and twenty-five percent for testing and verification. Anger, disgust, fear, happiness, neutral, sorrow, and surprise labels (1–7) were used to categorize each picture, with each of the four fractal aspects serving as an independent variable. In the end, the SVM approach was accurate enough to accurately identify around 93% of the photos tested. When it came to training models, DCNN was 100% accurate, whereas SVM was only 93% accurate. Comparing DCNN with SVM, the latter's test results accuracy was 89.33% while the former's was 98.8%. Psychologists assert that one's emotions play
a key part in providing care to patients. Therefore, we adopted a method to verify the whole idea with regards to efficacy, accuracy, and statistical analysis, considering these characteristics and experiments. For the sake of making headway in the treatment of illnesses like Parkinson's and Alzheimer's, we want to leverage evolutionary principles in our future research to develop facial expression categorization algorithms. Alzheimer's disease does not completely alter the facial appearance of its sufferers. Since our model requires processing and communication cycles in addition to evolutionary techniques, this might be a good challenge once again when applied to a diverse situation. Furthermore, we want to improve our approach to emotion identification by considering many angles of the face, the expression, and the posture.

Declarations

Funding: This study did not receive any funding.

Conflict of Interest: All authors declare that they have no conflicts of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References


50. C. A. Sen, “Ultrasound-guided transrectal implantation of fiducial markers for image-guided radiotherapy of prostate cancer without local anesthesia: Patient-reported gastrointestinal--
genitourinary system complications and pain,” 2020.


Figures

Figure 1

The conceptual diagram of the model
Figure 2

A description of a CNN architecture

Figure 3

Images samples in JEFFE dataset: images size is 256x256 grayscale image in tiff format
Figure 4

Feature extraction process over a plot of the fractal model

Figure 5

Fractal extracted functions in the form of the main image matrix
**Figure 6**

Confusion matrix of training data using SVM method
Figure 7

Confusion matrix of testing data using SVM method
Figure 8

The presented DCNN architecture

Figure 9
Fig 8. DCNN training process: accuracy and loss value for 300 iteration

Figure 10

Fig 9. The results of the confusion matrix for training data of DCNN
**Figure 11**

**Fig 10.** The results of the confusion matrix for testing data of DCNN

<table>
<thead>
<tr>
<th>Output Class</th>
<th>Anger</th>
<th>Disgust</th>
<th>Fear</th>
<th>Happy</th>
<th>Neutral</th>
<th>Sadness</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>90%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Disgust</td>
<td>0%</td>
<td>90%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Fear</td>
<td>0%</td>
<td>0%</td>
<td>90%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Happy</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>90%</td>
<td>0%</td>
<td>1%</td>
<td>0%</td>
</tr>
<tr>
<td>Neutral</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>90%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Sadness</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>89%</td>
<td>0%</td>
</tr>
<tr>
<td>Surprise</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>90%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target Class</th>
<th>Anger</th>
<th>Disgust</th>
<th>Fear</th>
<th>Happy</th>
<th>Neutral</th>
<th>Sadness</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Disgust</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>98.9%</td>
<td>1.1%</td>
</tr>
<tr>
<td>Fear</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Happy</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>98.9%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Neutral</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Sadness</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
<td>1.1%</td>
<td>0%</td>
</tr>
<tr>
<td>Surprise</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>99.8%</td>
</tr>
</tbody>
</table>

14.3%