Journal of Theoretical and Applied Information Technology
15th May 2025. Vol.103. No.9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

NOVEL ONE PROTOTYPE IN EACH OF THE CLASSES EMBEDDED IN FUZZY K-NEAREST NEIGHBOR

PAYUNGSAK KASEMSUMRAN[1] , TATSANEE CHAIYA[2]

[1] Department of Computer Science, Faculty of Science, Maejo University, Chiang Mai, Thailand

[2] School of Renewable Energy, Maejo University, Chiang Mai, Thailand

E-mail: [1] payungsak.kae@gmail.com, [2] tatsaneech@gmail.com

ABSTRACT

Many researchers in the world are interested in classification algorithms. One of the most famous and popular algorithms is K-NN and Fuzzy K-NN. Nevertheless, the time complexity of both algorithms is $O\left(n^2\right)$. In our previously publication, we introduced the novel algorithm in which each of the classes has one representation. Although it is a good algorithm and helps to solve a time process problem but there are still problems with the accuracy rate of classification. Thus, we develop a novel one prototype in each of classes embedded in Fuzzy K-Nearest Neighbor to solve the problem for accuracy rate of classification. The system provides 99.55%, 98.25%, 83.43%, 69.07%, 80.01%, 100%, 96.25%, 81.81%, 80.65% and 83.27%% in MIT-CBCL, ORL, FEI, Georgia Tech, Pain Expression, JAFFE, Senthikumar, Yale, PICS and CMU AMP databases, respectively.

Keywords: *Prototype, Embedded, Fuzzy K-NN, Representation, Novel algorithm, NsgFK-NN1P.*

1. INTRODUCTION

In recent days, many researchers around the world have shown interest in classification algorithms, such as K-NN [1, 2] and Fuzzy K-NN [3, 4]. These algorithms are among the most favored due to their simplicity and effectiveness in various applications, including image recognition, medical diagnosis, and pattern recognition. However, despite their popularity, these algorithms face significant challenges related to time complexity and classification accuracy, especially when dealing with large and diverse datasets.

Previous studies have primarily focused on improving the accuracy and efficiency of these algorithms. For instance, traditional K-NN algorithms classify data points based on the majority class among the k-nearest neighbors, while Fuzzy K-NN assigns membership values to each class, providing a more nuanced classification. However, these approaches often suffer from high computational costs, as they require calculating distances between the query point and all training samples, leading to increased time complexity.

Our work differs in motivation and findings by introducing a novel algorithm that embeds one prototype per class in Fuzzy K-Nearest Neighbor (FK-NN). This approach aims to enhance classification accuracy while significantly reducing time complexity. By using a single representative prototype for each class, our algorithm minimizes the number of distance calculations required, thereby improving computational efficiency without compromising accuracy.

Research Problem and Questions Despite the advancements in classification algorithms, there remains a significant gap in addressing the trade-off between accuracy and time complexity. This study aims to answer the following research questions:

How can the accuracy of classification algorithms be improved without increasing time complexity?

What is the impact of embedding one prototype per class in FK-NN on classification accuracy across diverse datasets?

How does the proposed algorithm compare to existing algorithms in terms of performance and efficiency?

Significance of the Study This study fills the gap in the literature by providing a comprehensive solution to the challenges of time complexity and classification accuracy. The proposed algorithm not only enhances the performance of FK-NN but also offers a practical solution for real-world applications where computational resources are limited. By addressing these challenges, our work contributes new knowledge to the field of classification algorithms

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and opens up new possibilities for their application in various domains, including healthcare, finance, and security.

Literature Review and Gaps Previous research has shown that while traditional K-NN and FK-NN algorithms are effective, they often struggle with high computational costs and reduced accuracy in certain scenarios. Studies such as those by Keller et al. [3] and Hunt et al. [4] have attempted to address these issues by incorporating fuzzy membership functions and other modifications. However, these approaches still fall short in terms of balancing accuracy and efficiency. Our novel algorithm builds on these foundations by introducing a single prototype per class, which not only reduces computational costs but also maintains high classification accuracy.

By addressing these gaps, our study provides a novel approach that enhances the performance of FK-NN algorithms, making them more suitable for a wide range of applications. This contribution is particularly significant in fields where quick and accurate classification is crucial, such as real-time image recognition and medical diagnostics.

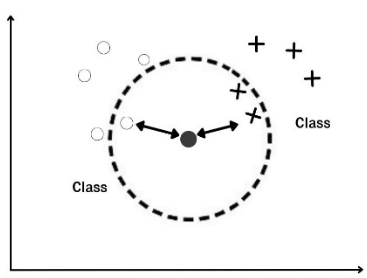

Figure 1 : Examples of a circle where it will be extended indefinitely until the three nearest samples are found.

We implemented two of our previous publication, one represents a new algorithm that can classify 2 signals with different amplitude but the same pattern into same class [5]. The second, new algorithm is shown hot to find one representative for the specific class from all data of the class [6]. These papers introduced new algorithm that could increase the accuracy rate of the original Fuzzy K-Nearest Neighbor but still have $O(n^2)$ time complexity. We used the same EEG dataset as Consumer Grade Brain Sensing for Emotion Recognition [7]. We will call the algorithm of second paper "NsgFK-NN1P Ver.1". After that, we produced an indirect comparison for comparing two of our algorithms. The second algorithm could reduce the time process of the first algorithm from $O(n^2)$ to $O(n)$. Because this algorithm uses the prototype which derived from the new algorithm or new equation and not the center. We are using these prototypes to replace the all samples for each of the classes. However, we found that the accuracy rate of the second algorithm decreased by about 10% from the first algorithm.

In this research, we introduce two new algorithms to solve this problem. After we experimented in the same dataset and the same environment. We found that the accuracy rate of the new first algorithm decreased but the new second algorithm increased while the time complexity still is $O(n)$.

Nevertheless, We compared two new algorithms and the previous algorithm [6] on ten standard face recognition datasets, i.e., ORL [8] or AT&T database, FEI face database [9], Yale [10], JAFFE (Japanese Female Facial Expression) database [11], Pain expressions database [12], Senthilkumar database [13], PICS [14], MIT-CBCL [15], CMU AMP database [16], Georgia Tech face database [17].

2. PROPOSED METHOD

2.1 System Description

We generated a new algorithm for increasing the accuracy rate when we use the one prototype per class for classification. We have imaginative brain signals. We used the same dataset from [5, 6, 7] which can describe as details below:

We put the Electroencephalography headset on forty-nine sample populations as shown in Figure 2. The signal has been generated as shown in Figure 3.

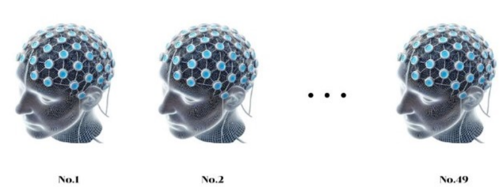

Figure 2 : Examples of 49 imaginative brain signals.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

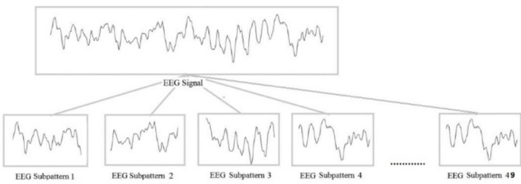

Figure 3 : Examples of 49 imaginative brain signals.

The signals of forty-nine sample populations have been asked to raise their left arms and right arms. The signal from right arms and left arms are generated into two classes as shown in Figure 4.

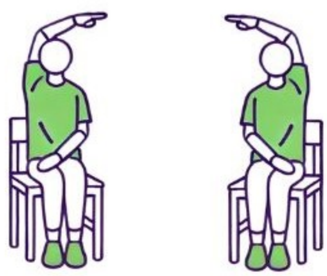

Figure 4 : A person who asked to raise their right arm and left arm.

We selected 9800 from the best high amplitude of a signal for each class. So, the summation of left and right arm signal is 19600 for the 2 classes as shown in Figure 5, Figure 6 and Figure 7, respectively.

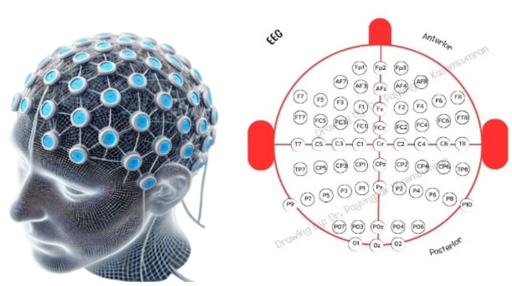

Figure 5 : Examples of 64 points or 64 channels on head.

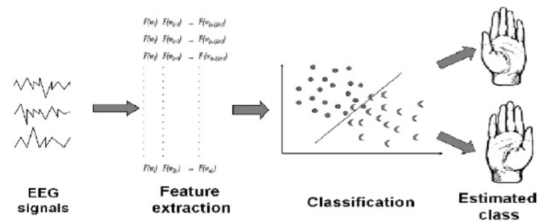

Figure 6 : Procedure of Novel One Prototype in Each of Classes Embedded in Fuzzy K-Nearest Neighbor Classification.

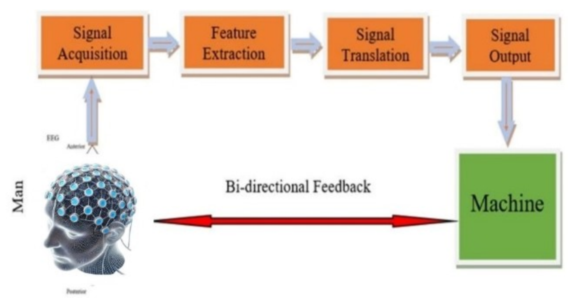

Figure 7 : Closed-loop of Novel One Prototype in Each of Classes Embedded in Fuzzy K-Nearest Neighbor architecture.

3. METHODOLOGY

A. This paper, we implemented the first new algorithm for data classification. We will call "NsgFK-NN1P Ver.2". The membership value modified from Vector Fuzzy C-Means (VFC) [18] as

$$u_i(\mathbf{x}) = \frac{\sum_{a=1}^{K} u_{ia}\left(1 \Big/ \left(\sum_{p=1}^{C} \frac{Lev(\mathbf{x}-\mathbf{x}_{med\,a}^{\;i})}{Lev(\mathbf{x}-\mathbf{x}_{med\,a}^{\;p})} \right)^{2/m-1} \right)}{\sum_{a=1}^{K} \left(1 \Big/ \left(\sum_{p=1}^{C} \frac{Lev(\mathbf{x}-\mathbf{x}_{med\,a}^{\;i})}{Lev(\mathbf{x}-\mathbf{x}_{med\,a}^{\;p})} \right)^{2/m-1} \right)} \; ; K \le C \qquad (1)$$

where

K : the number of training data nearest of testing data.

C : the number of classes when $K \le C$.

$\mathbf{x}_{meda}^{\;i}$: the median in class i for a = 1 to K.

Note that one constraint is that

$$\left(\sum_{p=1}^{C}\frac{Lev(\mathbf{x}-\mathbf{x}_{meda}^{\;i})}{Lev(\mathbf{x}-\mathbf{x}_{meda}^{\;p})}\right)^{2/m-1} \neq 0 \ ,$$

we can proof the properties of equation (1) as below.

$$\lim_{\left(\sum_{p=1}^{C}\frac{Lev(x-x_{meda}^{\;i})}{Lev(x-x_{meda}^{\;p})}\right)^{2/m-1}\to 0^{+}} \{u_i(\mathbf{x})\} = \lim_{\left(\sum_{p=1}^{C}\frac{Lev(x-x_{meda}^{\;i})}{Lev(x-x_{meda}^{\;p})}\right)^{2/m-1}\to 0^{+}} u_i(\mathbf{x})$$

$$\lim_{\left(\sum_{p=1}^{C}\frac{Lev(x-x_{meda}^{\;i})}{Lev(x-x_{meda}^{\;p})}\right)^{2/m-1}\to 0^{+}} u_i(\mathbf{x}) = \frac{\sum_{a=1}^{K} u_{ia}\left(1 \Big/ \left(\sum_{p=1}^{C}\frac{Lev(x-x_{meda}^{\;i})}{Lev(x-x_{meda}^{\;p})}\right)^{2/m-1}\right)}{\sum_{a=1}^{K}\left(1 \Big/ \left(\sum_{p=1}^{C}\frac{Lev(x-x_{meda}^{\;i})}{Lev(x-x_{meda}^{\;p})}\right)^{2/m-1}\right)} = 1 \quad \forall i,K \quad (2)$$

and

$$\lim_{\left(\sum_{p=1}^{C}\frac{Lev(x-x_{meda}^{\;i})}{Lev(x-x_{meda}^{\;p})}\right)^{2/m-1}\to \infty} \{u_i(\mathbf{x})\} = \lim_{\left(\sum_{p=1}^{C}\frac{Lev(x-x_{meda}^{\;i})}{Lev(x-x_{meda}^{\;p})}\right)^{2/m-1}\to \infty} u_i(\mathbf{x})$$

$$\lim_{\left(\sum_{p=1}^{C}\frac{Lev(x-x_{meda}^{\;i})}{Lev(x-x_{meda}^{\;p})}\right)^{2/m-1}\to \infty} u_i(\mathbf{x}) = \frac{\sum_{a=1}^{K} u_{ia}\left(1 \Big/ \left(\sum_{p=1}^{C}\frac{Lev(x-x_{meda}^{\;i})}{Lev(x-x_{meda}^{\;p})}\right)^{2/m-1}\right)}{\sum_{a=1}^{K}\left(1 \Big/ \left(\sum_{p=1}^{C}\frac{Lev(x-x_{meda}^{\;i})}{Lev(x-x_{meda}^{\;p})}\right)^{2/m-1}\right)} = 1 \quad \forall i,K \quad (3)$$

then $0 <= u_i(x) <= 1 \quad \forall i,K$

and $\sum_{i=1}^{c} u_i(x) = 1$ by experiment

From equations (2) and (3)

$$0 < \left(\sum_{p=1}^{C}\frac{Lev(x-x_{meda}^{\;i})}{Lev(x-x_{meda}^{\;p})}\right)^{2/m-1} <= \infty$$

and

$$\lim_{\left(\sum_{p=1}^{C}\frac{Lev(x-x_{meda}^{\;i})}{Lev(x-x_{meda}^{\;p})}\right)^{2/m-1}\to \infty} \{u_i(\mathbf{x})\} <= u_i(\mathbf{x}) = \frac{\sum_{a=1}^{K} u_{ia}\left(1 \Big/ \left(\sum_{p=1}^{C}\frac{Lev(x-x_{meda}^{\;i})}{Lev(x-x_{meda}^{\;p})}\right)^{2/m-1}\right)}{\sum_{a=1}^{K}\left(1 \Big/ \left(\sum_{p=1}^{C}\frac{Lev(x-x_{meda}^{\;i})}{Lev(x-x_{meda}^{\;p})}\right)^{2/m-1}\right)}$$

and

$$u_i(\mathbf{x}) = \frac{\sum_{a=1}^{K} u_{ia} \left(1 \Big/ \left(\sum_{p=1}^{C} \frac{Lev(x - x_{meda}^{\ i})}{Lev(x - x_{meda}^{\ p})} \right)^{2/m-1} \right)}{\sum_{a=1}^{K} \left(1 \Big/ \left(\sum_{p=1}^{C} \frac{Lev(x - x_{meda}^{\ i})}{Lev(x - x_{meda}^{\ p})} \right)^{2/m-1} \right)} <= \lim_{\left(\sum_{p=1}^{C} \frac{Lev(x - x_{meda}^{\ i})}{Lev(x - x_{meda}^{\ p})} \right)^{2/m-1} \to 0^+} \{u_i(\mathbf{x})\}$$

$$0 <= u_i(\mathbf{x}) = \frac{\sum_{a=1}^{K} u_{ia} \left(1 \Big/ \left(\sum_{p=1}^{C} \frac{Lev(x - x_{meda}^{\ i})}{Lev(x - x_{meda}^{\ p})} \right)^{2/m-1} \right)}{\sum_{a=1}^{K} \left(1 \Big/ \left(\sum_{p=1}^{C} \frac{Lev(x - x_{meda}^{\ i})}{Lev(x - x_{meda}^{\ p})} \right)^{2/m-1} \right)} <= 1$$

from $1 <= K <= n$, K is constant and

$\sum_{j=1}^{n} u_i(x) > 0$ where $i \in 1, \ldots, c$

then

$$0 < \sum_{j=1}^{n} u_i(\mathbf{x}) = \sum_{j=1}^{n} \frac{\sum_{a=1}^{K} u_{ia} \left(1 \Big/ \left(\sum_{p=1}^{C} \frac{Lev(x - x_{meda}^{\ i})}{Lev(x - x_{meda}^{\ p})} \right)^{2/m-1} \right)}{\sum_{a=1}^{K} \left(1 \Big/ \left(\sum_{p=1}^{C} \frac{Lev(x - x_{meda}^{\ i})}{Lev(x - x_{meda}^{\ p})} \right)^{2/m-1} \right)} <= n$$

or $\quad 0 < \sum_{j=1}^{n} u_i(x) <= n$

where $\quad u_i(x) = u_{ij} = u_i(x_j)$

then $\quad 0 < \sum_{j=1}^{n} u_i(x_j) <= n$

or $\quad 0 < \sum_{j=1}^{n} u_{ij} <= n$

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

and $\quad \sum_{i=1}^{c} u_i(x) = 1$

property of equation (1) is

$$u_{ij} = \left\{ u_{ij} \,\middle|\, 0 \le u_{ij} \le 1, \forall i = 1:c, j = 1:n; \forall i \exists j \right\}$$

when

$$u_{ij} > 0 \ and \ 0 < \sum_{j=1}^{n} u_{ij} \le n, \sum_{i}^{c} u_{ij} = 1 \ \forall j$$

i.e. The Computational complexity of Vector Fuzzy C-Means (VFC) is approximately $O(C \times K \times C \times N \log N) = O(N)$.

B. We implemented the second new algorithm for data classification which we call "Novel One Prototype in Each of Classes Embedded in Fuzzy K-Nearest Neighbor". For the membership value modified from New Fuzzy Entropy (NFE) [19] as

$$u_i(\mathbf{x}) = \frac{\sum_{a=1}^{K} u_{ia} \left(\sum_{p=1}^{C} e^{-\frac{1}{\gamma}\left(Lev\left(\mathbf{x} - \mathbf{x}_{meda}^{\ p}\right) - Lev\left(\mathbf{x} - \mathbf{x}_{meda}^{\ i}\right) \right)} \right)}{\sum_{a=1}^{K} \left(\sum_{p=1}^{C} e^{-\frac{1}{\gamma}\left(Lev\left(\mathbf{x} - \mathbf{x}_{meda}^{\ p}\right) - Lev\left(\mathbf{x} - \mathbf{x}_{meda}^{\ i}\right) \right)} \right)} ; K \le C \quad (4)$$

Note that one constraint is that

$$\sum_{p=1}^{C} e^{-\frac{1}{\gamma}\left(Lev\left(\mathbf{x} - \mathbf{x}_{meda}^{\ p}\right) - Lev\left(\mathbf{x} - \mathbf{x}_{meda}^{\ i}\right) \right)} \neq 0 ,$$

i.e. the divider cannot be zero.

Where $\gamma = 1$ (the degree of fuzziness) is a weighting exponent used for controlling the degree of fuzziness and the membership function same the FCM or used for controlling the compromise between the intra-cluster scattering error and the fuzzy entropy.

we can proof the properties of equation (4) as below.

$$\lim_{\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)\to 0}\{u_i(\mathbf{x})\} = \lim_{\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)\to 0} u_i(\mathbf{x})$$

$$\lim_{\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)\to 0} u_i(\mathbf{x}) = \frac{\sum_{a=1}^{K} u_{ia}\left(\sum_{p=1}^{C} e^{-\frac{1}{\gamma}\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)}\right)}{\sum_{a=1}^{K}\left(\sum_{p=1}^{C} e^{-\frac{1}{\gamma}\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)}\right)} = 1 \quad \forall i, K \quad (5)$$

and

$$\lim_{\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)\to \infty}\{u_i(\mathbf{x})\} = \lim_{\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)\to \infty} u_i(\mathbf{x})$$

$$\lim_{\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)\to \infty} u_i(\mathbf{x}) = \frac{\sum_{a=1}^{K} u_{ia}\left(\sum_{p=1}^{C} e^{-\frac{1}{\gamma}\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)}\right)}{\sum_{a=1}^{K}\left(\sum_{p=1}^{C} e^{-\frac{1}{\gamma}\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)}\right)}$$

$$\frac{\sum_{a=1}^{K} u_{ia}\left(\sum_{p=1}^{C} e^{-\frac{1}{\gamma}\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)}\right)}{\sum_{a=1}^{K}\left(\sum_{p=1}^{C} e^{-\frac{1}{\gamma}\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)}\right)} = 0 \quad \forall i, K \quad (6)$$

then $0 <= u_i(x) <= 1 \quad \forall i, K$

and $\sum_{i=1}^{c} u_i(x) = 1$ by experiment

From equations (5) and (6)

$$0 < \sum_{p=1}^{C} e^{-\frac{1}{\gamma}\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)} <= \infty$$

$$\lim_{\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)\to \infty}\{u_i(\mathbf{x})\} < u_i(\mathbf{x}) = \frac{\sum_{a=1}^{K} u_{ia}\left(\sum_{p=1}^{C} e^{-\frac{1}{\gamma}\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)}\right)}{\sum_{a=1}^{K}\left(\sum_{p=1}^{C} e^{-\frac{1}{\gamma}\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)}\right)}$$

and

$$u_i(\mathbf{x}) = \frac{\sum_{a=1}^{K} u_{ia}\left(\sum_{p=1}^{C} e^{-\frac{1}{\gamma}\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)}\right)}{\sum_{a=1}^{K}\left(\sum_{p=1}^{C} e^{-\frac{1}{\gamma}\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)}\right)} <= \lim_{\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)\to 0}\{u_i(\mathbf{x})\}$$

$$0 <= u_i(\mathbf{x}) = \frac{\sum_{a=1}^{K} u_{ia}\left(\sum_{p=1}^{C} e^{-\frac{1}{\gamma}\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)}\right)}{\sum_{a=1}^{K}\left(\sum_{p=1}^{C} e^{-\frac{1}{\gamma}\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)}\right)} <= 1$$

from $1 <= K <= n$; K is constant and

$\sum_{j=1}^{n} u_i(x) > 0$ where $i \in 1,\dots,c$

then

$$0 < \sum_{j=1}^{n} u_i(\mathbf{x}) = \sum_{j=1}^{n} \frac{\sum_{a=1}^{K} u_{ia}\left(\sum_{p=1}^{C} e^{-\frac{1}{\gamma}\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)}\right)}{\sum_{a=1}^{K}\left(\sum_{p=1}^{C} e^{-\frac{1}{\gamma}\left(Lev\left(x-x_{med_a}^{p}\right)-Lev\left(x-x_{med_a}^{i}\right)\right)}\right)} <= n$$

$$0 < \sum_{j=1}^{n} u_i(x) <= n$$

where $u_i(x) = u_{ij} = u_i(x_j)$

then $0 < \sum_{j=1}^{n} u_i(x_j) <= n$

or $0 < \sum_{j=1}^{n} u_{ij} <= n$

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

and from $\sum_{i=1}^{c} u_i(x) = 1$ the property of equation is

$$U_{ij} = \left\{ \begin{array}{l} u_{ij} \left| 0 \le u_{ij} \le 1, i = 1:c, j = 1:n; \forall i \exists j, \right. \\ u_{ij} > 0 \ and \ 0 < \sum_{j=1}^{n} u_{ij} \le n; \forall j, \sum_{i}^{c} u_{ij} = 1 \end{array} \right\}$$

The time complexity of New Fuzzy Entropy (NFE) or “Novel One Prototype in Each of Classes Embedded in Fuzzy K-Nearest Neighbor” is approximately $O(C \times K \times C \times N \log N) = O(N)$.

Table 1: The experimental results from NsgFK-NN1P Ver.1, NsgFK-NN1P Ver.2 and NsgFK-NN1P Ver.3 for EEG datasets [6].

Dataset	NsgFK-NN1P Ver.1	NsgFK-NN1P Ver.2	NsgFK-NN1P Ver.3
EEG	78.21%	72.77%	79.15%

Table 2: The experimental results from NsgFK-NN1P Ver.1, NsgFK-NN1P Ver.2 and NsgFK-NN1P Ver.3 for ten face recognition standard datasets.

Dataset	NsgFK-NN1P Ver.1	NsgFK-NN1P Ver.2	NsgFK-NN1P Ver.3	The other’ s
ORL	98.25%	88.47%	98.25%	98% [20]
MIT-CBCL	99.50%	99.50%	99.55%	92.7% [21]
Georgia Tech	66.93%	62.53%	69.07%	96.6% [22]
FEI	82.43%	76.71%	83.43%	96% [23]
JAFFE	100.00%	100.00%	100.00%	99.43%[24]
Pain Expression	79.05%	73.56%	80.01%	N/A
Senthilkumar	96.25%	96.25%	96.25%	83.33%[25]
PICS	79.69%	74.15%	80.65%	N/A
Yale	80.60%	83.03%	81.81%	86.67%[26]
CMU AMP	82.28%	76.57%	83.27%	90.5%[27]

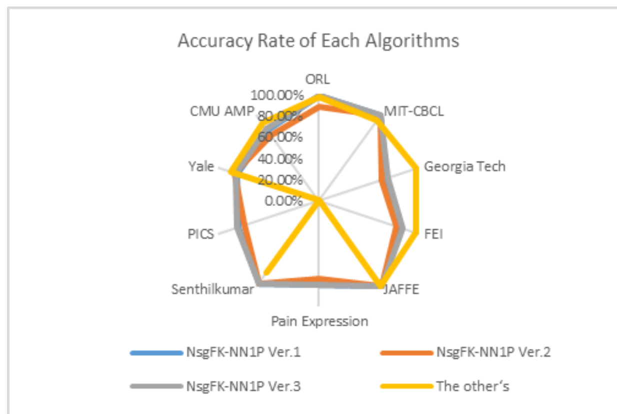

Figure 8 : Accuracy Rate of Each Algorithms on 2-D Column.

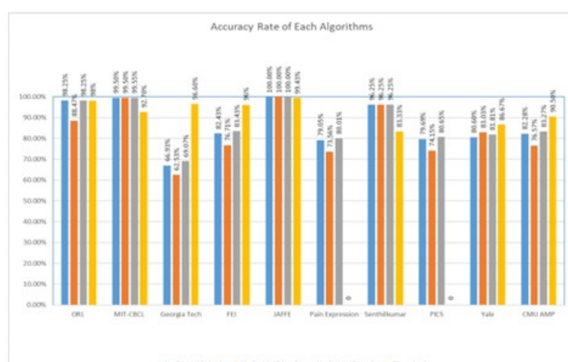

Figure 9 : Accuracy Rate of Each Algorithms on Radar.

Journal of Theoretical and Applied Information Technology
15th May 2025. Vol.103. No.9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

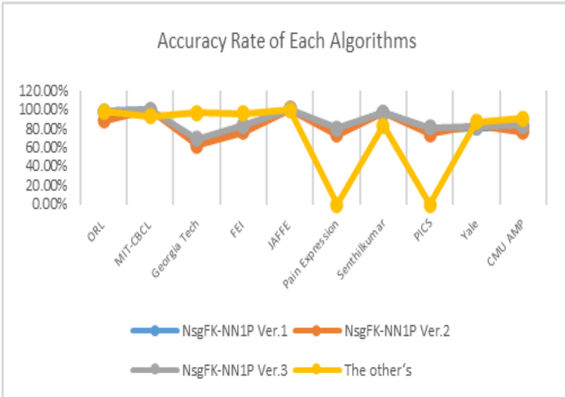

Figure 10 : Accuracy Rate of Each Algorithms on Radar.

Although, some dataset of the experiment has higher accuracy than our algorithm (NsgFK-NN1P Ver.3) but they have higher on time complexity. Some other algorithm on some datasets has $O(n^2)$ while in our algorithm has $O(n)$.

4. EXPERIMENTAL RESULTS

We implement NsgFK-NN1P Ver.1 in previous paper, NsgFK-NN1P Ver.2 and NsgFK-NN1P Ver.3 in this paper. We used the same dataset from [5, 6, 7] and We compared two new algorithms and the previous algorithm [6] on ten standard face recognition datasets, i.e., ORL [8] or AT&T database, FEI face database [9], Yale [10], JAFFE (Japanese Female Facial Expression) database [11], Pain expressions database [12], Senthilkumar database [13], PICS [14], MIT-CBCL [15], CMU AMP database [16], Georgia Tech face database [17].Two hundred and thirteen images from JAFFE database have been used in our research. These images contain variety of female facial expressions for one hundred and sixty-five gray scale images, 15 individuals from the Yale database in GIF format. Each subject has eleven images, which have a variety of characters and emotional, some of the sampling photos also had accessories like glasses.

From the results of these experiments mentioned above, the data correction shown to researchers that some datasets, the algorithm could provide the compare results with the existing algorithms and cropped pre-processing. For others datasets that came only face no background (sources from CMU AMP), the algorithm outperforms the existing algorithm significant. This study also found that the accuracy rate of NsgFK-NN1P Ver.1, NsgFK-NN1P Ver.2 and NsgFK-NN1P Ver.3 are 78.21%, 72.77% and 79.15%, respectively.

5. CONCLUSION

In a previous paper, we developed the Novel String Grammar Fuzzy K-Nearest Neighbor Techniques with One Prototype in Each Class. Although it is a good algorithm and helps to solve time process problem, there are still issues with the accuracy rate of classification. In this paper, we develop the Novel One Prototype in each Class Embedded in Fuzzy K-Nearest Neighbor to solve the problem for accuracy rate of classification. Nevertheless, this can reduce the time complexity of Fuzzy K-Nearest Neighbor Algorithm from down to. However, from the paper [28] and [29] which both papers have applied the fuzzy K-nearest neighbor algorithm to healthcare-related tasks. In the future, the authors intend to further apply this algorithm to other healthcare tasks or different fields as well.

While our novel algorithm significantly improves classification accuracy across various datasets, it is important to acknowledge certain limitations. For instance, the algorithm's performance may vary depending on the nature of the dataset and the specific characteristics of the data. Additionally, the reduction in time complexity, although substantial, may not be sufficient for real-time applications in certain scenarios. Future research should focus on addressing these limitations and exploring potential enhancements to further optimize the algorithm's performance.

Research Contributions This study contributes to the field of classification algorithms by introducing a novel approach that balances accuracy and time complexity. The proposed algorithm provides new insights into the use of prototypes in FK-NN, offering a practical solution to improve classification performance. Furthermore, the findings highlight the potential of this algorithm in various applications, including healthcare-related tasks, thereby expanding its significance in the area.

Open Research Issues Despite the improvements achieved by our novel algorithm, several open research issues remain. These include:

Exploring the algorithm's applicability to real-time applications and its performance in dynamic environments.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Investigating the impact of different types of datasets on the algorithm's accuracy and efficiency.

Developing strategies to further reduce time complexity while maintaining high classification accuracy. In this paper, we develop the Novel One Prototype in each Class Embedded in Fuzzy K-Nearest Neighbor to solve the problem for accuracy rate of classification. Nevertheless, this can reduce the time complexity of Fuzzy K-Nearest Neighbor Algorithm from $O(n^2)$ down to $O(n)$.

REFERENCES:

[1] E. Fix and J.L. Hodges, "Discriminatory analysis, nonparametric discrimination: Consistency properties", Technical Report 4, USAF School of Aviation Medicine, 1951.

[2] C. Elkan, "Nearest Neighbor Classification", January 11, 2011.

[3] J. M. Keller, M. R. Gray and J.A. Givens Jr., "A Fuzzy K-Nearest Neighbor Algorithm", IEEE Transactions on systems, man, and cybernetics, Vol. SMC-15, No. 4, pp. 580 – 585, 1985.

[4] J. M. Keller and D. J. Hunt, "Incorporating Fuzzy Membership Functions into the Perceptron Algorithm", IEEE Transactions on pattern analysis and machine intelligence, Vol. PAMI-7, No. 6, November 1985.

[5] P. Kasemsumran, and E. Boonchieng, "EEG-Based Motor Imagery Classification Using Novel Adaptive Threshold Feature Extraction and String Grammar Fuzzy K-Nearest Neighbor Classification," Journal of Computers, Vol. 30 No. 2, pp. 27-40, 2019.

[6] P. Kasemsumran and E. Boonchieng, "EEG-Based Motor Imagery Classification Using Novel String Grammar Fuzzy K-Nearest Neighbor Techniques with One Prototype in Each of Classes", International Conference on Artificial Intelligence in Information and Communication (ICAIIC), doi: 10.1109/ICAIIC48513.2020.9065236. pp. 742-745, Fukuoka, Japan, 2020,

[7] P. Lakhan, N. Banluesombatkul, V. Changniam, R. Dhithijaiyratn, P. Leelaarporn, E. Boonchieng, and S. Hompoonsup, "Consumer Grade Brain Sensing for Emotion Recogni-tion", IEEE Sensors Journal, Vol. 19, Issue: 21, 2019.

[8] ORL Face Database, AT&T Laboratories Cambridge, 1992-1994.

[9] FEI face database, http://fei.edu.br/~cet/facedatabase.html

[10] A. Georghiades, Yale face database, Center for computational Vision and Control at Yale University, http://cvc.cs.yale.edu/cvc/projects/yalefaces/yalefaces.html, 2002.

[11] Japanese Female Facial Expression (JAFFE) Database, http://www.kasrl.org/jaffe.html.

[12] Pain expressions, http://pics.psych.stir.ac.uk/2D_face_sets.htm.

[13] Senthilkumar database, http://www.geocities.ws/senthilirtt/Senthil%20Face%20Database%20Version1.

[14] PICS - Psychological Image Collection at Stirling, http://pics.psych.stir.ac.uk/

[15] Wang, Y., "Face Recognition by Nonsubsmpled ContoURLet Transform and Support Vector Machine", in Z. Zhong (Ed.), Proceedings of the International Conference on Information Engineering and Applications (IEA), Vol. 1, pp. 595- 600, 2012.

[16] Advanced Multimedia Processing Lab http://chenlab.ece.cornell.edu/projects/FaceAuthentication/download.html

[17] Georgia Tech Face Database, ftp://ftp.ee.gatech.edu/pub/users/hayes/facedb/.

[18] S. Ganguly, D. Bose and A. Konar, "Clustering using Vector Membership: An Extension of the Fuzzy C-Means Algorithm", Fifth International Conference on Advanced Computing (ICoAC), 2013.

[19] D. R. Yang, L. S. Lan, and W. C. Pao, "A New Fuzzy Entropy Clustering Method with Controllable Membership Characteristics", MWSCAS' 06. 49th IEEE International Midwest Symposium on, Aug 6-9, 2006.

[20] M. H. Yang, N. Ahuja and D. Kriegman, "Face Recognition Using Kernel EigenFaces", 2000 International Conference on Image Processing, pp. 37 – 40, 2000.

[21] Y. Wang, "Face Recognition by Nonsubsmpled ContoURLet Transform and Support Vector Machine", in Z. Zhong (Ed.), Proceedings of the International Conference on Information Engineering and Applications (IEA), Vol. 1, pp. 595- 600, 2012.

[22] L. Chen, H. Man and A.V. Nefian., "Face recognition based on multi-class mapping of fisher scores," Pattern Recognition, Vol. 38, No. 6. pp. 799–811, 2005.

[23] B. Mohamed and S. Mohamed, “Face Recognition using Two Dimensional Subspace Analysis and PNN”, International Journal of Computer Applications, Vol. 72, No. 6, May 2013.

[24] J. Zhang, S. Z. Li, and J. Wang, “Manifold Learning and Applications in Recognition”, in Y-P. Tan, K. H. Yap, and L. Wang (Eds.), Intelligent Multimedia Processing with Soft Compu-ting, pp. 281- 300, 2005.

[25] G. V. Sugar, S. Y. Barker, K. B. Raja and K. R. Venugopal, “Convolution based Face Recognition using DWT and Feature Vector Compression”, Third International Conference on Image Infonnation Processing, 2015.

[26] S. Valarmathy, K. M. Arun and R. Sangeetha, “Evaluation of Face Recognition Using Vector Features In Local Pattern Descriptors”, Third International Conference on Devices, Circuits and Systems (ICDCS'16), 2016.

[27] K. C. Ting, D. B. L. Bong and Y. C. Wang, “Performance Analysis of Single and Combined Bit-Planes Feature Extraction for Recognition in Face Expression Database”, Proceedings of the International Conference on Computer and Communication Engineering, May 13-15, 2008.

[28] Q. Zhang, J. Sheng, Q. Zhang, L. Wang, Z. Yang and Y. Xin, “Enhanced Harris hawks optimization-based fuzzy k-nearest neighbor algorithm for diagnosis of Alzheimer’s disease”, Computers in Biology and Medicine, https://doi.org/10.1016/j.compbiomed.2023.107392, (https://www.sciencedirect.com/science/article/pii/S0010482523008570)107392, ISSN 0010-4825, Vol. 165, 2023.

[29] J. Xing, C. Li, P. Wu, X. Cai and J. Ouyang, “Optimized fuzzy K-nearest neighbor approach for accurate lung cancer prediction based on radial endobronchial ultrasonography”, Computers in Biology and Medicine, 108038, https://doi.org/10.1016/j.compbiomed.2024.108038, https://www.sciencedirect.com/science/article/pii/S0010482524001227, ISSN 0010-4825, Vol. 171, 2024.