

Fuzzy Classifier with Convolution for Classification of Handwritten Digits

Phạm Minh Tâm¹

¹⁻² Viện công nghệ thông tin & Truyền thông, Trường Đại học Bách Khoa Hà Nội,

Giảng viên hướng dẫn Phạm Văn Hải

TÓM TẮT— Phân loại bằng logic mờ là một trong những lĩnh vực quan trọng của trí tuệ nhân tạo. Mô hình phân loại dựa trên những hàm kích hoạt mờ để phân loại những bộ dữ liệu không chắc chắn hoặc có ngữ nghĩa. Mô hình phân loại mờ yêu cầu kiến thức biết trước, và luật mờ quá đơn giản để đạt được độ chính xác cao trên những bộ dữ liệu phức tạp như chữ số viết tay và các dữ liệu hình ảnh. Bộ phân loại được đề xuất trong bài có thể chia làm hai phần: trích xuất đặc trưng từ lớp đầu ra của mạng CNN và hàm kích hoạt mờ Guass để phân loại dữ liệu. Logic mờ và mạng neural là hai lĩnh vực rất quan trọng và được ứng dụng rộng rãi trong các hệ thống trí tuệ nhân tạo và trợ giúp quyết định. Sự kết hợp giữa hai phương pháp này có thể giúp cải thiện kết quả. Sử dụng thuật toán học máy gradient descent, những thông số trên mô hình sẽ được huấn luyện phù hợp với dữ liệu. Mô hình có thể tự học để trích xuất ra những đặc trưng phù hợp cho việc phân loại. Thử nghiệm thực hiện trên bộ MNIST. Với bộ dữ liệu này, mô hình mạng phân lớp mờ với CNN đạt được kết quả cao.

Từ khóa— Deep learning, fuzzy classification, fuzzy rule, machine learning.

I. GIỚI THIỆU

Phân loại hình ảnh là một vấn đề quan trọng, có nhiều nghiên cứu, bài báo được thực hiện để giải quyết vấn đề này ví dụ mạng CNN [1,10]. Trong thực tế, rất nhiều hình ảnh có nhiễu và tín hiệu ngẫu nhiên khác khiến cho việc phân loại hình ảnh gặp nhiều khó khăn.

Bộ phân loại mờ là một mô hình trí tuệ nhân tạo sử dụng hàm kích hoạt mờ. Bộ phân lớp tính toán xác suất của một mẫu dữ liệu phụ thuộc vào một nhãn nào đó bằng cách sử dụng hàm kích hoạt mờ. Phân lớp mờ có nhiều ưu điểm trên bộ dữ liệu biến động và có nhiễu. Vì vậy các bộ phân lớp mờ thường được dùng trong xử lý ảnh [16]. Tuy nhiên các phương pháp phân lớp mờ cũ gặp nhiều khó khăn trong việc trích xuất đặc trưng từ ảnh. Để có đặc trưng, cần phải có hiểu biết [3, 14], vì vậy làm tăng lượng công việc cần xử lý bởi con người. Đồng thời, các phương pháp phân loại mờ cũ khó khăn trong việc đạt độ chính xác cao, cụ thể như việc nhận dạng chữ viết tay trên bộ MNIST.

Để tận dụng những ưu thế của bộ phân lớp mờ trong việc xử lý dữ liệu biến động, nhiều nghiên cứu cố gắng kết hợp những mô hình học sâu với thuật toán xử lý mờ. Trong số các mô hình học sâu, thuật toán CNN đạt được nhiều ưu điểm vượt trội cho bài toán phân loại hình ảnh [11]. Các đặc trưng của ảnh sẽ tự động trích xuất thông qua các lớp convolution. Do đó, để cải thiện kết quả của bộ phân loại mờ, các đặc trưng được trích xuất từ lớp tích chập của mạng CNN sẽ trở thành đầu vào cho bộ phân lớp mờ. Các lớp tích chập đóng vai trò tiền xử lý để đưa ra đặc trưng cho bộ phân lớp mờ. Kết hợp mạng neural và logic mờ để giải quyết bài toán dự đoán chữ số viết tay trong ảnh.

Bài trình bày theo cấu trúc sau: chương 2 trình bày cơ sở lý thuyết, chương 3 giới thiệu mô hình đề xuất, chương 4 là thực nghiệm và cuối cùng chương 5 là kết luận.

II. Cơ sở lý thuyết

II.1 Tích chập

Convolutional Neural Network (CNNs – Mạng nơ-ron tích chập) là một trong những mô hình Deep Learning tiên tiến giúp cho chúng ta xây dựng được những hệ thống thông minh với độ chính xác cao như hiện nay như hệ thống nhận diện khuôn mặt, hệ thống xe tự lái. Mạng CNN sử dụng phép toán tích chập (convolutional) để tính toán đưa ra đặc trưng của dữ liệu đầu vào.

Mạng CNN bao gồm một tập các lớp CNN xếp chồng lên nhau. Dữ liệu (thường là hình ảnh) được đưa qua các lớp. Sau mỗi lớp tích chập, dữ liệu tăng dần tính trừu tượng và thể hiện tính đặc trưng cao hơn so với dữ liệu trước đó.

Phép toán tích chập được sử dụng trong mạng CNN.

Định nghĩa phép tích chập với hàm liên tục như sau:

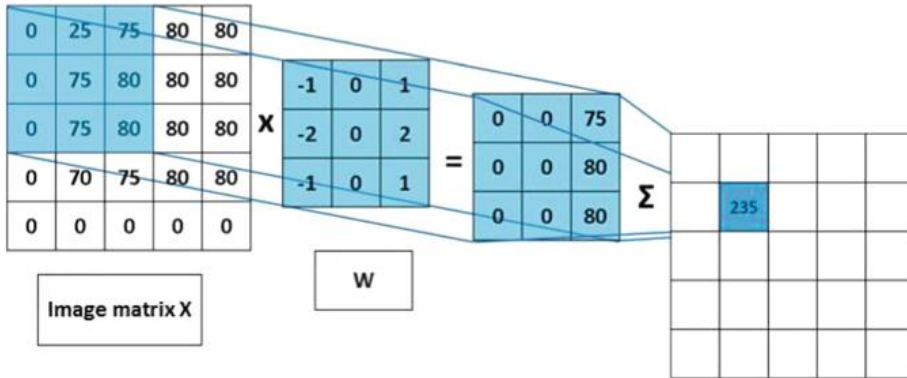
$$y(x) = \int_{-\infty}^{\infty} f(\tau)g(x - \tau)d\tau \quad (1)$$

$f(x)$ và $g(x)$ là hàm khả tích trong miền xác định. Phép toán tích chập được sử dụng rộng rãi trong xử lý tín hiệu và các lĩnh vực khác. Đối với hàm rời rạc

$$y(n) = \sum_{i=-\infty}^{\infty} x(i)h(n - i) = x(n) * h(n) \quad (2)$$

Với dữ liệu hình ảnh như bộ MNIST, hình ảnh chữ viết tay là một mảng hai chiều, có thể được coi như một tín hiệu rời rạc.

Phép toán tích chập trên ảnh chữ viết tay sử dụng ma trận nhân chập W trượt trên ma trận dữ liệu ảnh X . Các giá trị trên ma trận dữ liệu là mức độ xám của các điểm trên ảnh. Ma trận W trượt trên toàn bộ ảnh, từ đó giúp trích xuất các đặc trưng của ảnh. Các đặc trưng khác nhau có thể được trích xuất từ những ma trận nhân chập khác nhau.



Hình 1: Phép toán tích chập

II.2 Hàm kích hoạt mờ

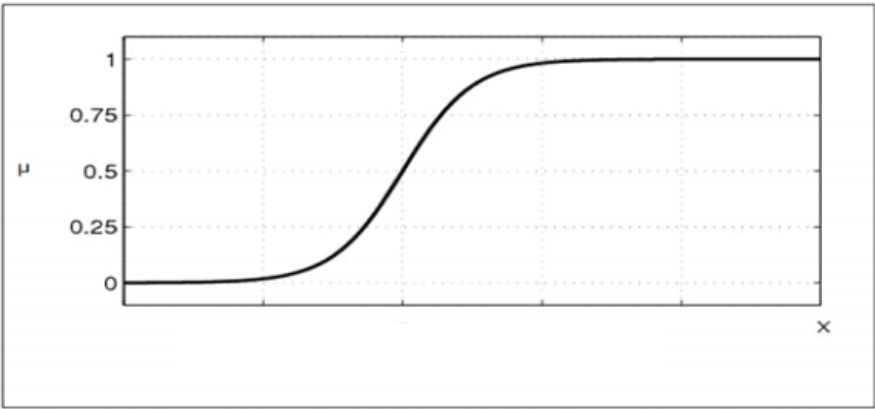
Hàm kích hoạt mờ là hàm tính toán độ mờ của một giá trị trong tập mờ. Trong logic mờ, nó đại diện cho mức độ thuộc của một phần tử đến tập mờ tương ứng. Độ thuộc thường bị nhầm lẫn với xác suất, mặc dù chúng khác biệt về mặt khái niệm, bởi vì độ thuộc mờ thể hiện tư cách thành viên trong các tập mờ, không có khả năng xảy ra sự kiện hoặc điều kiện nào đó [18].

Cho tập mờ A , $A = \{(x, \mu_A(x)) | x \in X\}$ với $\mu_A(x)$ là độ thuộc của phần tử x đối với tập A được tính bằng hàm kích hoạt mờ. Độ thuộc của phần tử x nằm trong khoảng $[0,1]$.

II.2.1 Hàm kích hoạt mờ Sigmoid

Hàm thành viên sigmoidal có hai tham số: chịu trách nhiệm về độ dốc của nó tại điểm giao nhau $x = c$ [18]. Hàm thành viên của hàm sigmoid có thể được biểu diễn dưới dạng $\text{Sigmf}(x; a, c)$ và nó là

$$\text{sigmf}(x; a, c) = \frac{1}{1 + e^{-a(x-c)}} \tag{3}$$

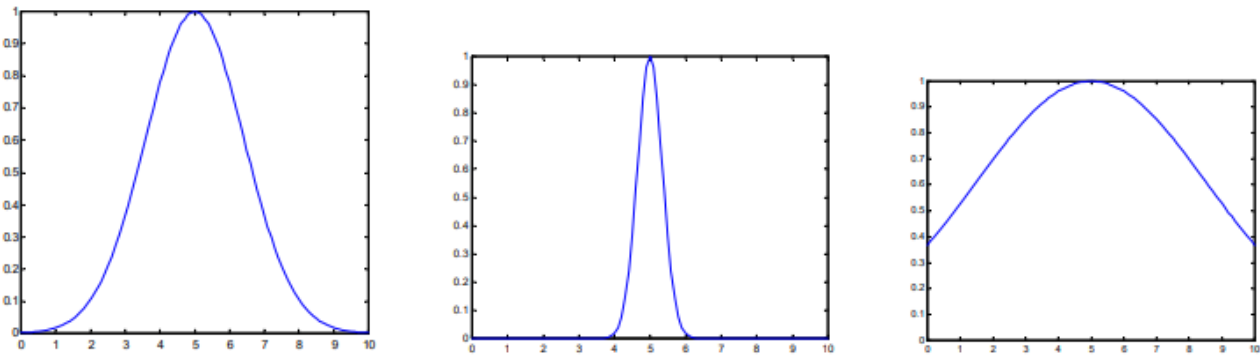


Hình 2: Hàm sigmoid

II.2.2 Hàm kích hoạt mờ Gauss

Phân phối gauss cho thấy rằng nếu một biến ngẫu nhiên x tuân theo phân phối Gauss với kỳ vọng x_0 và độ lệch chuẩn σ_0 được biểu thị bằng $x \sim N(x_0, \sigma_0^2)$, thì hàm mật độ xác suất của nó được tính theo công thức:

$$f(x) = ae^{-(x-x_0)^2/2\sigma_0^2} \tag{4}$$

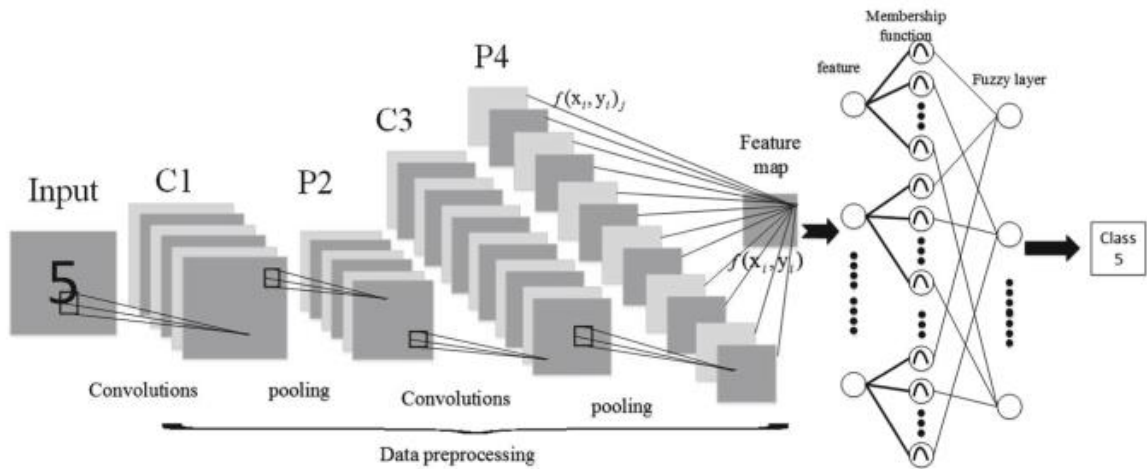


Hình 3: Hàm gauss với $(x_0, \sigma_0) = (5,0.5)$, $(x_0, \sigma_0) = (5,2)$ và $(x_0, \sigma_0) = (5,5)$

Kỳ vọng toán học x_0 xác định vị trí trung tâm của phân phối gauss. Tham số a xác định biên độ của hàm phân phối. Hàm Gauss là hàm kích hoạt mờ thường được sử dụng trong mô hình phân loại mờ [7,12]. Kết quả của hàm có thể thấy rằng x càng gần x_0 , giá trị càng lớn. Do đó, trong mô hình phân loại mờ, x_0 thường được sử dụng làm giá trị mốc để phân lớp đối một mẫu đối tượng. Đối với một mẫu mới, giá trị tương ứng càng lớn, nó càng gần với giá trị mốc x_0 của lớp đó. Và khả năng đối tượng thuộc về lớp đó cần.

III. Mô hình đề xuất

Trình phân loại được đề xuất kết hợp phép toán tích chập với bộ phân loại mờ để thực hiện phân loại bộ dữ liệu chữ viết tay MNIST. Trong số đó, phần phép toán tích chập có thể được coi là quá trình tiền xử lý dữ liệu, bao gồm các phép toán tích chập, phép toán pooling, để trích xuất các đặc trưng của dữ liệu đầu vào. Phần phân loại mờ Gauss tính toán phân loại các lớp của từng mẫu dữ liệu đầu vào và đưa ra quyết định. Trong quá trình xây dựng mô hình, các tham số của lớp tích chập, giá trị trung bình và phương sai của hàm kích hoạt mờ Gauss được tối ưu hóa bằng thuật toán lan truyền ngược.



Hình 4: Mô hình phân lớp kết hợp CNN với logic mờ

III.1 Lớp trích xuất đặc trưng

Quá trình trích xuất đặc trưng của dữ liệu sử dụng mạng CNN bao gồm các lớp chập C1 và C3, các lớp gộp P2 và P4 và một lớp tổng hợp các đặc trưng. Các lớp tích chập sử dụng ma trận kernel kích thước (3x3). Phép toán tích chập được sử dụng để trích xuất các đặc trưng cục bộ của hình ảnh chữ số viết tay và phép toán gộp có thể làm giảm kích thước của các đặc trưng trong khi vẫn giữ được thông tin hữu ích. Các phép toán tích chập và gộp bảo tồn thông tin vị trí hình ảnh trong khi vẫn trích xuất các đặc trưng của ảnh đó. Giá trị đặc trưng có thể thu được bằng cách thêm lớp tổng hợp có trọng số vào sau lớp P4. Ví dụ: nếu mẫu đầu vào là hình ảnh màu xám, đầu vào là ma trận hai chiều có kích thước là $M * N$. Sau khi khởi tạo ngẫu nhiên các ma trận chập k , k đặc trưng có thể thu được sau lớp tích chập C1. Và kích thước của C1 là $M * N$. Sau đó đi qua P2 thu được k ma trận kích thước $(M/2) (N/2)$. Vì vậy, sau hai lớp tích chập và hai lớp gộp, hình ảnh được chuyển thành tập các đặc trưng. Sau đó, để có được đặc trưng tổng quát của toàn bộ hình ảnh, ta đưa các đặc trưng này qua một lớp tổng hợp được tính toán theo công thức:

$$f(x_i, y_i) = \sum_j w_j f(x_i, y_i)_j \quad (5)$$

Trong đó w_j là trọng số

$f(x_i, y_i)_j$ là giá trị tại điểm (x_i, y_i) của ma trận đặc trưng thứ j của lớp P2

III.2 Lớp phân loại mờ

Đối với mô hình phân loại mờ, cần xác định hàm kích hoạt của từng loại. Đối với một mẫu đầu vào, ta sẽ tính toán độ thuộc của mẫu đó đối với các lớp. Lớp có độ thuộc lớn nhất sẽ là lớp đại

diện của mẫu đó. Như được trình bày ở trên, đặc trưng của ảnh đầu vào là một ma trận hai chiều. Ta cần chuyển ma trận hai chiều đó thành một vector theo công thức:

$$f_k(k \in [0, n]), f_k = f(x_i, y_i) \quad (6)$$

với $k = i+j$

Đối với một mô hình phân loại gồm C lớp, tham số được khởi tạo ngẫu nhiên. Và độ thuộc của một mẫu với một lớp được tính toán. Độ thuộc các lớp của một mẫu sẽ được tổng hợp lại thành một vector, sau đó lớp có độ thuộc lớn nhất sẽ trở thành lớp đại diện cho mẫu đó. Công thức tính toán được hiển thị trong biểu thức:

$$pro_J = \sum_I e^{-(f_I - \mu_{IJ})^2 / 2\sigma_{IJ}^2} \quad (7)$$

$$J = \operatorname{argmax}_I \left(\sum e^{-(f_I - \mu_{IJ})^2 / 2\sigma_{IJ}^2} \right) \quad (8)$$

với f_i là đặc trưng của dữ liệu mà ta thu được khi đi qua mạng CNN. μ_{ij} là điểm mốc của lớp j . σ là độ lệch chuẩn. Lớp có độ thuộc cao nhất sẽ được chọn là lớp của dữ liệu.

III.3 Huấn luyện

Các tham số mô hình đều được khởi tạo ngẫu nhiên. Để tăng độ chính xác của mô hình cần phải huấn luyện thông qua hàm mất mát. Hàm mất mát được định nghĩa:

$$L = - \sum_J y_{lab_J} \times \log(pro_J) \quad (9)$$

với $proj$ là kết quả phân lớp của mô hình, y_{labj} là nhãn biết trước.

Huấn luyện mô hình khi hàm mất mát đạt đủ nhỏ:

$$\min(L) \leq \varepsilon \quad (10)$$

Với ε là ngưỡng hàm loss.

IV. KẾT QUẢ THỬ NGHIỆM

Mô hình được sử dụng để phân loại chữ số viết tay với bộ dữ liệu MNIST và bộ dữ liệu Fashion MNIST.

MNIST là bộ dữ liệu chữ số từ 0-9 được viết bằng tay bởi nhân viên của U.S. Census Bureau và các học sinh trung học. Bộ dữ liệu bao gồm 60000 ảnh để huấn luyện và 10000 ảnh để kiểm tra. Các hình chữ số đều được đưa về cùng một kích thước là 28×28 điểm ảnh. Mỗi điểm có giá trị từ 0-255 mô tả cường độ sáng của điểm đó.

Mô hình sử dụng trên bộ dữ liệu FashionMNIST, là bộ dữ liệu gồm 60000 ảnh để huấn luyện và 10000 ảnh để kiểm tra. Các hình chữ số đều được đưa về cùng một kích thước là 28×28 điểm ảnh. Mỗi điểm có giá trị từ 0-255 mô tả cường độ sáng của điểm đó. FashionMNIST gồm ảnh thuộc 10 loại trang phục.

Dữ liệu trước khi đưa vào mô hình được chuẩn hóa về trong khoảng (0,1) bằng cách chia tất cả các giá trị cho 255.

Sử dụng thư viện keras để xây dựng mô hình. Thời gian huấn luyện khoảng 12 phút, epochs = 20.

Với hàm kích hoạt mờ gauss:

Dataset	Accuracy
MNIST	98,15%
Fashion MNIST	88,58%

Bảng 1: Kết mô hình CNN-fuzzy sử dụng hàm gauss

Với hàm kích hoạt mờ sigmoid:

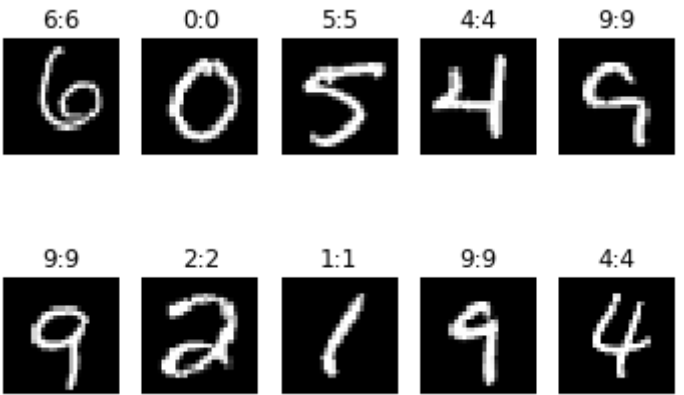
Dataset	Accuracy
MNIST	95,25%
Fashion MNIST	82,82%

Bảng 2: Kết mô hình CNN-fuzzy sử dụng hàm sigmoid

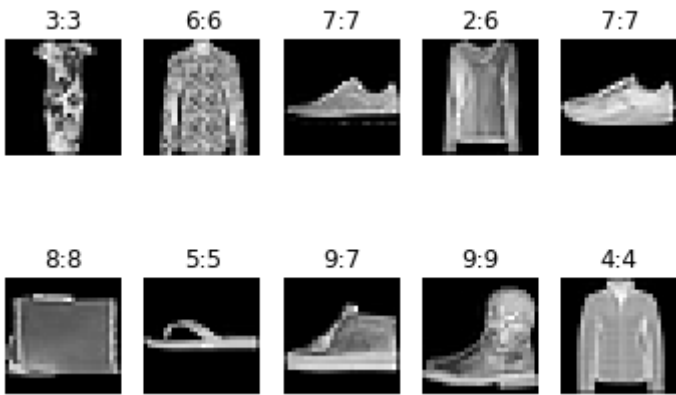
Với mô hình CNN không sử dụng hàm mờ

Dataset	Accuracy
MNIST	97,73%
Fashion MNIST	88,10%

Bảng 3: Mô hình CNN không dùng mờ



Hình 5: Kết quả mô hình mở chạy trên bộ chữ số viết tay MNIST



Hình 6: Kết quả mô hình mở chạy trên bộ FashionMNIST

V. KẾT LUẬN

Mô hình phân loại này kết hợp mô hình tích chập CNN với phân loại mờ để phân loại các chữ số viết tay. Các lớp tích chập được sử dụng để trích xuất các đặc trưng trong hình ảnh chữ số viết tay. Bằng cách sử dụng thuật toán lan truyền ngược, phương pháp học có giám sát có thể được sử dụng để tối ưu hóa các tham số, giúp mô hình phân loại mờ được đề xuất có thể đạt được độ chính xác mong muốn. Mô hình phân loại mờ sẽ học để trích xuất các đặc trưng cần thiết của ảnh đầu vào cho phù hợp và đưa ra kết quả tốt nhất. Mô hình tự học này làm giảm khối lượng công việc của con người, máy tính có thể tự động chạy. Đồng thời, kết quả thử nghiệm trên bộ dữ liệu của MNIST cho thấy so với mô hình phân loại truyền thống, độ chính xác phân loại của mô hình phân loại đã được cải thiện rất nhiều.

VI. TÀI LIỆU THAM KHẢO

- [1] Alex, K., Ilya, S., Geoffrey, E.H.: Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25, 1097–1105 (2012)
- [2] Arthur, D., Vassilvitskii, S.: k-means++: The advantages of careful seeding. In: *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1027–1035. Society for Industrial and Applied Mathematics (2007)
- [3] Chuang, K.S., Tzeng, H.L., Chen, S., Wu, J., Chen, T.J.: Fuzzy c-means clustering with spatial information for image segmentation. *Comput. Med. Imaging Graph.* 30(1), 9–15 (2006)
- [4] Duan, X., Wang, Y., Pedrycz, W., Liu, X., Wang, C., Li, Z.: AFSNN: a classification algorithm using axiomatic fuzzy sets and neural networks. *IEEE Trans. Fuzzy Syst.* 26(5), 3151–3163 (2018)
- [5] Deng, Y., Ren, Z., Kong, Y., Bao, F., Dai, Q.: A hierarchical fused fuzzy deep neural network for data classification. *IEEE Trans. Fuzzy Syst.* 25(4), 1006–1012 (2017)
- [6] Fan, H.W., Zhang, G.Y., Ding, A.L., Xie, C.R., Xu, T.: Improved BP algorithm and its application in detection of pavement crack. *J. Chang'an Univ.* 30(1), 438–457 (2010)
- [7] Hameed, I.A.: Using gaussian membership functions for improving the reliability and robustness of students' evaluation systems. *Expert Syst. Appl.* 38(6), 7135–7142 (2011)
- [8] Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
- [9] Kulkarni, A.D., Lulla, K.: Fuzzy neural network models for supervised classification: multispectral image analysis. *Geocarto Int.* 14(4), 42–51 (1999)
- [10] Lawrence, S., Giles, C.L., Tsoi, A.C., Back, A.D.: Face recognition: a convolutional neural-network approach. *IEEE Trans. Neural Netw.* 8(1), 98–113 (2002)
- [11] Li, H., Lin, Z., Shen, X., Brandt, J., Hua, G.: A convolutional neural network cascade for face detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5325–5334 (2015)
- [12] Tay, K.M., Lim, C.P.: Optimization of Gaussian fuzzy membership functions and evaluation of the monotonicity property of fuzzy inference systems. In: *IEEE International Conference on Fuzzy Systems*, pp. 1219–1224 (2011)
- [13] Van der Wilk, M., Rasmussen, C.E., Hensman, J.: Convolutional Gaussian processes. In: *Advances in Neural Information Processing Systems*, pp. 2849–2858 (2017)
- [14] Winkler, R., Klawonn, F., Kruse, R.: Fuzzy c-means in high dimensional spaces. *Int. J. Fuzzy Syst. Appl. (IJFSA)* 1(1), 1–16 (2011)
- [15] Xiaodong, D., Zedong, L., Cunrui, W., Back, A.D.: Research on multi-ethnic face semantic description and mining method based on AFS. *Chin. J. Comput.* 39, 1435–1449 (2016)

[16] Yongchuan, T., Yunsong, X.: Learning disjunctive concepts based on fuzzy semantic cell models through principles of justifiable granularity and maximum fuzzy entropy. *Knowl.-Based Syst.* 161, 268–293 (2018)

[17] Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: *European Conference on Computer Vision*, pp. 818–833 (2014)

[18] Fuzzy Membership Function Formulation and Parameterization

<https://cse.iitkgp.ac.in/~dsamanta/courses/archive/sca/Archives/Chapter%203%20Fuzzy%20Membership%20Functions.pdf>