ISSN 1840-4855
e-ISSN 2233-0046

Original scientific article
http://dx.doi.org/10.70102/afts.2025.1833.556

AN EARLY DETECTION OF KAWASAKI DISEASE USING ENHANCED FEATURE SELECTION AND HYBRIDIZATION-BASED CLASSIFICATION APPROACH

C. Kavitha[1], Dr.A. Subramani[2]

[1]Research Scholar, Mother Teresa Women's University, Kodaikanal, Dindigul, Tamil Nadu, India. e-mail: kavithaphd2021@gmail.com, orcid: https://orcid.org/0009-0008-8783-3676
[2]Assistant Professor, Dept. of Computer Science, M.V. Muthiah Govt. Arts College for Women, Dindigul, Tamil Nadu, India. e-mail: subramani.appavu@gmail.com, orcid: https://orcid.org/0000-0002-8303-8770

Received: June 25, 2025; Revised: September 09, 2025; Accepted: October 06, 2025; Published: October 30, 2025

SUMMARY

Kawasaki disease (KD) is a disease that cause inflammation in blood vessels, impacting kids and causes heart and blood vessel damage. Patients with KD who have medium or big aneurysms, developing coronary dilatation for a minimum of 2 months, are related to the worst late coronary results, which are not widely researched. The primary goal of the suggested technique is to introduce an intelligent framework that can perform clinical data mining more accurately. In this, clinical data mining is done by using the novel hybridized classification approach. Initially, to replace the missing data, the Random Forest (RF) algorithm is proposed. Then, using the optimization algorithm, the feature selection is done, namely the artificial bee colony approach, which would select the best feature from the training data set. F-score values are considered as the fitness values for optimal feature selection. Finally, the classification process is done using the proposed hybridization approach; clustering is done before classification and in every iteration of classification. The clustering is done using the FCM clustering, and Deep neural network is employed for classification. The simulation results identified that the proposed hybridization approach has better classification accuracy, leading to an efficient assessment of KD.

Key words: *kawasaki disease (kd), data imputation, random forest, feature selection, artificial bee colony algorithm, hybridization-based classification, deep neural network, fuzzy c-means clustering.*

INTRODUCTION

The nonspecific systemic vascular inflammation known as Kawasaki disease (KD) occurs during infancy and early childhood. KD has been the primary reason for acquired cardiac illness in children, with an increasing frequency over time [1]. According to the American Heart Association, except for therapy, more than a quarter of patients with KD will have coronary artery aneurysms. Adults under the age of 40, they account for 5% of all acute coronary syndromes. Those who are unresponsive to (Intravenous immunoglobulin) IVIG have a high risk of coronary vein lesions, which was around 15% to 20%. According to previous research, initial treatment customized to each patient's risk of developing KD can enhance their prognosis. Therefore, early IVIG-resistance prediction can play an essential role in helping healthcare professionals' decision-making, implying the need for adequate risk assessment

methods to be developed. Machine learning (ML) is used in the autoimmune disease field for (i) Classification of patients with systemic sclerosis, (ii) systemic lupus erythematosus with the risk of patients (SLE). In addition, ML has been used to predict medication repurposing in immune-mediated cutaneous diseases to further our understanding of the pathogenesis of rheumatic disorders. A shortage of important scientific findings has been documented about the pathogenic processes of SAIDs [4]. It was specially employed in two studies employing SNPs to recognize IVIG resistance in patients with KD and distinguish those at a greater impact of forming coronary artery abnormalities (CAAs) using ML [2].

Electronic health records (EHRs) help to understand the personalized health problems, treatment responses, and clinical outcomes. For clinical occurrences, EHRs data develop diagnostic and predictive models [5]. For example, data from EHRs, such as longitudinal laboratory tests and clinical notes, were combined to create effective models for predicting how quickly chronic kidney disease will progress [21]. EHRs can be used to recognize whether a patient is at risk of stroke within a year based on their medical history and examination results. Disadvantages of randomized clinical trials, which consist of confounding factors, unaccounted-for comorbidities, and small sample sizes, can be solved via EHR-based clinical advances. When used as a foundation for intelligent systems, EHRs can help researchers discover new clinical evidence, make it easier to apply what they've learned in the clinic, and enhance the delivery of individualized healthcare [3].

Missing data, on the other hand, is a real-world issue that frequently arises in scientific settings. Several statistical analyses necessitate the inclusion of all available data, making missing data a serious issue. A statistical analysis that demands all data must make this choice for researchers who want to utilize it [7]. Inferential power is reduced, and critical information may be lost if missing data is thrown away. When dealing with high-dimensional and large-scale datasets like genomic, proteomic and neuroimaging, many of the statistical techniques that have been established so far are ineffective [8]. Most experts agree that multiple imputations should include all of your variables, despite the differences in your estimates of correlations.

It is critical for pattern recognition and ML to have a good feature selection and extraction mechanism in place. The calculation cost is minimized while the categorization accuracy is enhanced with the feature selection method [19]. Machine learning and data mining challenges are complicated by describing data from all features uniformly [9]. For classification and regression problems, the feature selection procedure should be employed to boost classification performance and lower the computational cost of the classifier. To optimize a criteria function's value overall M subsets of size, a segment of M features is selected within a collection of N attributes, where M<N, using a technique known as feature selection [10]. Essentially, a categorization operation is a tool used to categorize patterns or to assess the effectiveness of every subset in estimating the class output or a pattern.

Human cognition is the goal of artificial intelligence (AI). Health care is being transformed due to an explosion in healthcare data and quick advancements in analytics approaches [17]. There are many similarities between convolutional neural networks and standard ANNs in that they have neurons that self-optimize through learning. Still, every neuron obtains an input (multiplication followed by a nonlinear function) and operates (a scalar product). The whole network is formed as a single perceptive scoring function (the weight) from raw input image vectors to the final output image with corresponding score [11]. Automated KD diagnosis has recently been made possible by the development of many CNN models. However, the classic CNN algorithms may produce erroneous predictions when dealing with noisy databases. When the data is handled directly, classification becomes more complex.

Section II briefly describes the recent techniques for the prediction of KD; Section III describes the proposed methodology. Section IV represents accuracy analysis of proposed method. Lastly, section V depicts conclusion and future work.

REVIEWS OF LITERATURE

Many conventional methods using classic data mining and soft computing techniques are published here. In addition, this section examines the role and usefulness of various supervised learning and nature-

inspired diagnostic approaches for KD. The publication trends of relevant papers have also been studied from several angles. These techniques will be used in the future for diagnosing psychological illnesses.

Tremoulet et al. [12] designed a panel of biomarkers to differentiate acute KD-affected patients compared to febrile controls (FC) with better efficiency. There were three separate groups of KD and FC patients who were identified within the initial ten days of fever. Plasma samples were taken from each of these three groups. Luminex bead technology was used to measure the levels of 88 biomarkers linked to infection. Patients with KD were correctly diagnosed by biomarkers (platelet count, fibrinogen concentrations, glutamyl transferase, serum globulin, alanine transaminase, glutamyl transferase, sed rate, and absolute neutrophil count) in hospital laboratories. The eight-biomarker panel could be used to recognize KD correctly.

Sosa et al. [13] presented a detection approach for assumed individuals of partial KD. However, the diagnosis of certain children, particularly new born under 6 months and teenagers, is extremely difficult. For both complete and incomplete KD the treatment is the same. The mainstays of treatment are intravenous immunoglobulin and acetylsalicylic acid.

Burns et al. [14] developed a case-control approach over 17 years (2002-2019) to analyse 47 clusters obtained from 1332 patients with KD from a clinical lab. Synthetic KD clusters-based comparison is carried out for the cluster characteristics. Our study identified a "true" Kawasaki disease group consisting of five patients within a 7-day motion window. Univariate & multivariate empirical orthogonal function analyses evaluated KD clusters regarding clinical and demographic characteristics and average values for common laboratory data. Many actual KD cluster median values surpassed 95% compared to the two synthetic clusters in the univariate study.

Pezoulas et al. [15] developed an approach to identify important genes employed as diagnostic biomarkers for KD in affected patients along with the same gene expression profiles across the Convalescent phases, Acute, and Subacute using Self-Organizing Maps (SOMs). Then, utilizing a false discovery rate (FDR)-based attribute collection approach, genes with significant deviations between the per-phase clusters were identified and further analyzed. In this study, we discovered five genes that could be used to diagnose KD, including HLA-DQB1, ZBTB48, TNFRSF13C, and CASD1. This is the first time that these five genes have ever been published in the scientific literature. Here, the influence of the newly found genes was proven by training boosting ensembles for KD categorization on datasets using AdaBoost and XGBoost. The classifiers achieved an average of 4.40%, 5.52%, and 3.57% in accuracy, sensitivity, and specificity, followed by an improvement of accuracy by 2.30 %, sensitivity by 2.20% and specificity by 4.70%.

Kuniyoshi et al. [16] designed an ideal ML technique for identifying IVIG resistance in children with KD. Children with KD were analyzed by the medical records of 98 children (2–109 months old) admitted to the hospital. We discovered that 20% of the youngsters were initially resistant to IVIG therapy. We used linear SVM, logistic regression, and extreme gradient boosting to train three ML approaches against IVIG resistance. Furthermore, we used a layered 5-fold cross-validation to determine the predictive power (CV). We also used the recursive feature reduction technique to choose variables and run the nested 5-fold CV, similarly using certain variables. Regardless of how well the ML models predicted the future, we compared them to the present system. The specializations were greater than or equal to 0.90, but the sensitivities were lower than in previous scoring systems.

Manlhiot et al. [27] developed an etiologic structure including 1) proportion of the susceptible population, 2) modulation of risk, identified by chronic vulnerability to ecological risks, modifications of biological units, atmospheric, & contiguity with transferrable illnesses; and 3) weak to presumed spark [23]. Eventually, modeling of personal risk and global distribution is enhanced by considering non-traditional attributes.

Ling et al. [18] designed a diagnostic approach to assist physicians in differentiating KD patients from febrile controls to start the treatment. To further differentiate patients with Kawasaki sickness from healthy controls, researchers used a combination of clinical multivariate analysis, whole blood cell type-

specific gene expression studies, and urine peptidome profiling. An independent cohort of 30 KD and 30 control urine peptidomes showed an area under the receiver operating characteristic curve (ROC) of 0.919 for 139 putative indicators. Thirteen of these markers were verified in a further 30 samples of each disease and control group. A multivariate analysis of seven clinical characteristics with a ROC AUC of 0.803 and the inclusion of urine/blood-based biomarker panels efficiently classified 441 KD patients and 342 febrile control individuals.

Liu et al. [28] developed an effective approach for recognizing KD. A tertiary medical hospital in Taiwan enrolled 169 KD patients from 2009–2013. All patients' records were reviewed. After the acute KD period, echocardiography was conducted for various time frames. Using logistic regression, the danger variables for coronary aneurysms and increasing coronary dilatation were evaluated [32]. Out of 169 KD patients, 31 (18.3%) had maximum coronary Z-scores of +2.5 during the acute phase, 16 (9.5%) developed coronary aneurysms one month later, and 5 (3.1%) exhibited progressive coronary dilatation. Coronary aneurysm development one month after KD commencement was not significantly related to intravenous immunoglobulin non-responsiveness (P = 0.058) [6].

Chu et al. [20] developed an approach using Logistic regression analysis and BP neural network, respectively. The two approaches' diagnostic results were compared. Results: The study included 905 KD patients and 438 other feverish illnesses: 1 042 patients (700 KD, 342 other febrile illnesses) as the training group and 301 patients (205 KD, 96 other feverish illnesses) as the testing cohort. Univariate analysis revealed 37 factors that distinguished KD from other feverish illnesses. The ideal regression equation had 16 variables, according to logistic regression. This BP neural network has 37 inputs, 24 hidden, and 1 output node. The ROC analysis using Logistic regression revealed in the training cohort with an AUC of 0.91 and 0.89 in the testing cohort. The AUC of the BP neural network was 0.94 and 0.92, and the accuracy was 96.4% and 86%. The BP neural network model had more specificity than the Logistic regression approach. Therefore, the BP neural network model created offers significant supplementary diagnostic utility for KD diagnosis [25].

Antoon, J. W et al. [29] designed an approach to distinguish KD from other feverish diseases. They hypothesized that such an approach might be used in information technology to help physicians detect KD. Children with KD for the training phase = 276; children had KD-free fever (n=243); and for testing phase =136); and FC for the testing phase (n=121).

Wang et al. [22] developed a technique to utilize the multi-source data based on CNN and fusion models. Comparing suggested method's accuracy (AUC of 0.97) to many benchmark approaches. This strategy is used to enhance clinical data mining. In the course of our research, we were able to show the usefulness of matrices-based feature representation as well as CNN-based attribute extraction for partial data mining.

The most present research is insightful but has small sample sizes and uneven risk variables. The lack of sensitive risk indicators and effective techniques restricts these studies' ability to predict IVIG-resistant patients properly. Traditional algorithms cannot handle noisy databases properly, resulting in erroneous predictions. In addition, direct data handling makes classification more challenging. The suggested research framework's major goal is to develop an intelligent framework for clinical data mining.

PROPOSED METHODOLOGY

The research model's primary goal is to develop an intelligent framework capable of more precisely performing clinical data mining.

Firstly, the random forest approach for missing data imputation was suggested. RF appears to have all the qualities needed to solve missing information issues.

To pick features, to find the most optimal ones in training data sets, a method known as the Artificial bee colony technique is used. To determine the best features, the f-score value is used. Sort N features

with their F-Scores in descending order, and then compute each feature's F-Score. One or more features can be used to create the feature subset.

Using the suggested hybridization method, categorization is completed with clustering as a pre-requisite and subsequent iterations. FCM clustering and DNN classification are used in this instance for clustering and categorization, respectively. Figure 1. illustrates the process of the proposed methodology.

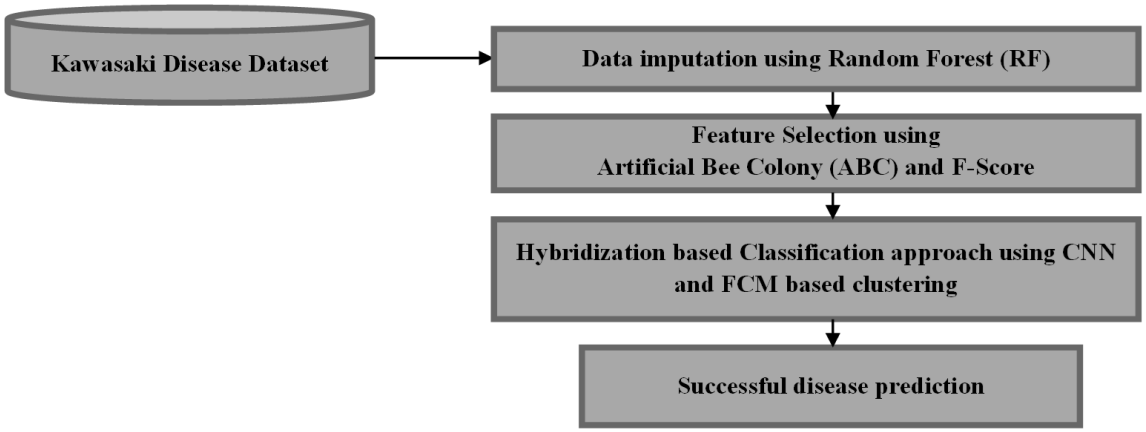

Figure 1. The process of the proposed methodology

Data Imputation

RFs [30] are one promising method. RF has the desirable properties of (1) managing mixed sorts of missing data; (2) dealing with interconnections and irregularity; (3) expanding to large proportions for reducing overfitting; and (4) yielding important variables metrics useful for variable selection. For RF missing data imputation, three general techniques have been used:

1. Preassign the data; expand the forest; based on the data's proximity adjust the initial missing values. Then, improve your results by iterating.

2. Concurrently estimate data as the forest grows; repeat for better outcomes.

3. Preassign the information; create a forest using every missing value; forecast the missing values. Finally, improve your results by iterating.

RF missing data techniques are now available in a variety of forms. The original RF proximity approach is briefly detailed in the section below.

Proximity imputation: RF_{prx} and RF_{prxR}

In this paper, proximity imputation is presented for filling in missing data in a KD dataset. The data is initially crudely imputed using strawman imputation in this approach. The median of non-missing values is used to estimate hidden values for continuous variables, and the most commonly occurring non-hidden value is employed for missing categorical variables. This estimated data is used to suit an RF. The (i, j) row captures the in-bag frequency that i and j occupy the similar terminal node. Using the proximity matrix, the actual missing values are assigned. The highest average proximity across non-hidden information is utilized or categorical variables. The modified information is utilized to create a novel RF, which is then executed.

Pure random splitting is an extreme example of random splitting. By picking a variable, the tree node is split and a split-point at random; splitting criteria are not applied. Using pure random splitting, We also apply RF_{prx}, and is represented as RF_{prxR}. And implement iterated versions of RF_{prx} and RF_{prxR}.

Feature Selection Using Artificial Bee Colony (ABC) Algorithm

Attribute sets that are substantially connected with the class, but are not correlated with other attributes in the set, are sought. Here, ABC method was utilized to eliminate class imbalance. Honey bees' foraging behaviour can be modelled using a new optimization technique that was just proposed. A honey bee swarm is a group of bees that work together to complete a task [24]. Employed bees, onlookers, and scout bees make up the three bee kinds in the ABC algorithm. Bees employed by a hive use their memory to locate nearby food sources, and they pass along this information to other bees in the hive as they do so. These observers tend to pick out the best food sources among those located by the employed bees. Onlooker bees are more likely to select a better quality (fitness) food source than lower quality. Bees that have been employed to look for new food sources are used to create scout bees. Data from the logical graph behaviour profile (diversity coefficient and state transition probability) is given to the feature selection method to improve credit card fraud detection.

The ABC generates an initial population of SN solutions (food sources) that are randomly dispersed, where SN is a substitute for the size of the swarm. With the assumption that $Xi = xi,1,xi,2,\ldots,xi,D$, the *i*th solution is denoted, where *D* is the dimension size. Each working bee X produces a new potential solution around its current location in the following way:

$$v_{i,j} = x_{i,j} + \Phi_{i,j}.(x_{i,j} - x_{k,j}) \quad (1)$$

The variable ak represents a candidate solution that was picked at random ($i = k$), b stands for a random dimension index that was chosen from the set {1, 2, D}, and ϕb is a random value between -1 and 1. Keep Xi unchanged if its fitness value is larger than that of Vi; otherwise, update Vi with Xi. By waggling their waggle dances, the employed bees alert the spectator bees of the food sources they've found. Observers select a feeding source based on its nectar abundance. A roulette wheel selection process is used to make the below probabilistic selection:

$$P_i = \frac{fit_i}{\sum_{j=1}^{SN} fit_j} \quad (2)$$

In which fit*i* denotes the fitness value of the *i*th solution. For this, better the result i , the more likely the food source is to be chosen. Food sources are abandoned when they can't be enhanced over predetermined (called limit) cycles. The scout bee develops a fresh food source to replace Xi if the abandoned source is Xi:

$$x_{i,j} = lb_j + rand(0,1).(ub_j - lb_j) \quad (3)$$

The variable rand (0, 1) is a distribute random number between 0 and 1, also the variables lb and ub represent the bottom and upper limits of the bth dimension, respectively.

Feature Score (F-Score)

It is measured by estimating the distinction of two groups of absolute numbers. Let training vectors x_k, k=1, 2, m, hen F-score is expressed as,

$$F(i) = \frac{(\bar{x}_i^{(+)} - \bar{x}_i)^2 + (\bar{x}_i^{(-)} - \bar{x}_i)^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n+}(x_{k,i}^{(+)} - \bar{x}_i^{(+)})^2 + \frac{1}{n_- - 1}\sum_{k=1}^{n-}(x_{k,i}^{(+)} - \bar{x}_i^{(-)})^2} \quad (4)$$

In which $\bar{x}_i$, $\bar{x}_i^{(+)}, \bar{x}_i^{(-)}$ represent the mean of ith attribute of the whole, negative, and positive datasets. The ith attribute of the kth positive samples is $x_{k,i}^{(+)}$ and the ith attribute of the kth negative samples is $x_{k,i}^{(-)}$.

Algorithm 1. Feature selection using Artificial Bee Colony (ABC) Algorithm

Input: Raw data

Output: Optimal features-based data

Assign the set of cards holders' data $x_i, i = 1,2, \ldots . . SN$.

Assess each $x_i, i = 1,2, \ldots . . SN$.

If the maximum number of iterations has not been achieved or if a "good enough" solution has not been identified, then

For i=1 to SN do

Create $u_i \; with \; x_i$

Assess u_i

iffit$(u_i) \geq fit(x_i)$then

$$x_i = u_i$$

for i=1 to SN do

Choose an employed bee

Attempt to enhance food source quality

Create a novel random food source

Store the optimal food source attained till now;

Upgrade the fitness value using F-score and Sort N numbers of feature's using Eq.(4)

Upgrade the novel solution

End

Hybridization Based Classification Approach

To conclude, the suggested hybridization strategy is used to classify data utilizing clustering before and after each classification iteration. FCM clustering and Deep neural network classification are used in this instance for clustering and classification, respectively.

Fuzzy C Means Clustering:

Images can be separated using the fuzzy c-mean technique, which divides the image into various cluster regions with comparable pixel values [31]. Here, a data item can be assigned a degree of membership in a cluster based on its degree of membership. By reducing the weighted within-group sum of squared error goal function, an iterative clustering approach creates an ideal c separation.

We will assume that c is greater than 1 and that the dataset is $X = \{x\_1, \ldots \ldots, x\_n\}$. For every i from 1 to c, the indicator function $\mu\_1, \ldots \ldots \mu\_c$ is defined as follows: $\mu\_i \; (x) = 1 \; if \; x$ is in Xi and $\mu\_i \; (x) = 0$ if x is not in Xi. In other words, mutually disjoint sets indicate a partition of X into c clusters $X\_1, \ldots . , X\_c$ such that $X\_1 \cup \ldots . . \cup \; X\_c = X$. For any x in X, the sum of all values of $\mu\_i(x)$ from 0 to 1 is equal to 1 thanks to a fuzzy extension that allows $\mu\_i(x)$ to take values on the range [0,1]. Here, $\{\mu\_1, \ldots \ldots \mu\_c\}$ represents a fuzzy c-partition of X. Therefore, JFCM may be written as

$$J_{FCM}(\mu, v) = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^{m} d^2(x_j, v_i) \qquad (5)$$

where $\mu = \{\mu_1, \ldots \ldots \mu_c\}$ is a fuzzy c-partition with $\mu_{ij} = \mu_i(x_j)$, the weighted exponent m is a fixed number larger than 1, $v = \{v_1, \ldots \ldots v_c\}$ is the c cluster centers, and $d^2(x_j, v_i) = \left\|x_j - v_i\right\|^2$ represents the Euclidean distance. The JFCM is upgraded with the below equations:

$$v_i = \frac{\sum_{j=1}^{n} \mu_{ij}^{m} x_j}{\sum_{j=1}^{n} \mu_{ij}^{m}} \ (i = 1, \dots \dots c) \qquad (6)$$

and

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} (\frac{d(x_j, v_i)}{d(x_j, v_i)})^{2(m-1)}} \qquad (7)$$

At each iteration, μ and v are upgraded using (6) and (7). Here $JFCM(\mu, v)$ is optimized until $|\mu(l+1) - \mu^l| \leq \varepsilon$ is the number of iterations.

For example, in (5), it is evident that FCM's objective function considers mainly each image pixel as an isolated point. In addition, d2 (x j,v i) evaluates the resemblance between the pixel intensity and the cluster center, determines the membership function in (7). Therefore, cluster membership has a greater value if the intensity values are closer to the cluster center.

CNN:

It uses convolution, ReLU, fully-connected layers, and pooling layers. Each connection in a network is a parameter in a typical ANN, where each layer is connected to every other layer. In addition, a local connection between neurons is used. The outcomes of receptive fields and neurons utilizing similar kernels are recorded in the activation map. Weight sharing is the term used to describe CNNs' ability to share their weight. Therefore, the activation maps generated by different kernels will differ, and the sum of kernels is changed using hyper-parameters. Because the weights are only related to the kernel's size, no matter how many connections presents in a network between neurons, they are only related to the kernel's size.

A convolutional layer is a necessary component in constructing a CNN, and it aims to transform the input into a depiction of a more existential level [26]. The convolutional layer performs operations between the input and the neurons using local connectivity, without using the full connectivity. The following are vital characteristics:

1. Kernel Size (N): Every kernel does have a window size, also known as a receptive field. A convolution operation is performed on an area from the input that matches with its window size, and the results will be stored in a map.

2. Stride (S): It specifies the number of pixels transferred to the newest location by the kernel. If set to 1, then each kernel will perform convolution operations before shifting one pixel until it joins the given input edge. As a result, the stride reduces the size of the activation maps.

3. Zero-padding (P): It defines how many zeros should be padded around the input's border. This is highly useful for retaining the input's dimension.

Shared weights and local connections reduce the network's overall number of parameters. For example, a convolutional layer with two kernels and size 4 of a local receptive field has a dimension of 4 x 4 x 3 with an input of 100 x 100 x 3. Because all 100 neurons in the layer share similar weights for every kernel, there will be 4 × 4 × 3 × 2 = 96 parameters per 100 neurons. Because local values in images are connected, and the analysis created by the local values are unchangeable in position, local convolutional processes in images result in results that contain specific image features. To extract patterns from an image, use a kernel with the same weights across all local regions. to the convolutional integers between the kernel and the input, non-linear activation function (ReLu, tanh, sigmoid, etc.) is commonly employed.

Back Propagation Algorithm

Convolution and sampling are used in CNN. The deconvolution of the input image is done with a trainable filter Fx for obtaining Cx's convolution layer. Sx + 1 is a narrow n time feature map created by pooling pixels from each neighbourhood, adding scalar weighting Wx + 1 weighted, and lastly, activation. The CNN utilizes a local receptive field, weight sharing, and subsampling for minimizing training attributes.

$$O_{x,y}^{(l,k)} = tanh \sum_{t=0}^{f-1} \sum_{r=0}^{K_h} \sum_{c=0}^{K_w} W_{(r,c)}^{(k,t)} O_{(x+r,x+c)}^{(l-1,t)} + Bias^{(l,k)} \quad (8)$$

The output of the neuron and the kth feature pattern are represented by f. Row x and column y in the lth sub-sample layer.

$$O_{x,y}^{(l,k)} = \tanh(W^k \sum_{r=0}^{S_h} \sum_{c=0}^{S_w} O_{(x\times S_h+r,y\times S_w+c)}^{(l-1,t)} + Bias^{(l,k)} \quad (9)$$

Output is given by:

$$O_{(i,j)} = \tanh\ (W^k \sum_{k=0}^{S-1} \sum_{x=0}^{S_h} \sum_{y=0}^{S_w} W_{(x,y)}^{(j,k)} O_{(x,y)}^{(l-1,t)} + Bias^{(l,k)} \quad (10)$$

s is the patterns in the sample layer. the output of the ith neuron l th output layer F

$$O_{(i,j)} = \tanh\ (\sum_{j=0}^{H} O_{(l-1,j)} W_{(i,j)}^{l} + Bias^{(l,i)} \quad (11)$$

A pooling layer is arranged amidst each successive layer. By preserving as much information as feasible, using a pre-specified pooling mechanism, pooling layers reduce the input's size. In addition, adding spatial invariance to the network via a pooling layer can further improve the model's generalizability. You can reduce the input dimension by half, for example, by setting the stride to 2, increasing the window size to 2, and using a zeros-padding of 0.

A completely linked layer is always placed amidst the penultimate and the output layers for typical CNN design to describe the attribute values' non-linear relationships. However, the above arrangement results in overfitting. So, the researchers have been adopting approaches like max-over time pooling to build CNN architectures without the need for a completely linked layer.

In the following part, we'll discuss how the strategies above can accurately forecast the onset of Kawasaki Disease.

RESULTS AND DISCUSSION

These findings are compared with those of previous algorithms for the identification of KD in this section. As a result of this new research, there is optimism that these people will be detected earlier and more effectively. According to equations (12)– (15), statistical measurements are also used to evaluate the average results for the classifiers.

Precision is the rate of accurately detected positive findings to predicted positive findings.

$$\text{Precision} = TP/\ FP+\ TP \quad (12)$$

A recall is the rate of accurately determined positive findings to findings in a real class.

$$\text{Recall} = TP/\ FN + TP \quad (13)$$

F1 measure is the mean of Precision & Recall, which consider false negatives & false positives.

$$\text{F1 measure} = 2*(\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall} \quad (14)$$

Accuracy is detected by Equation (15)

$$\text{Accuracy} = (FP+ TP)/(TN+TP+FN+FP) \quad (15)$$

In this context, TP, FP, TN, and FN signifies True Positive, False Positive, True Negative, and False Negative, respectively.

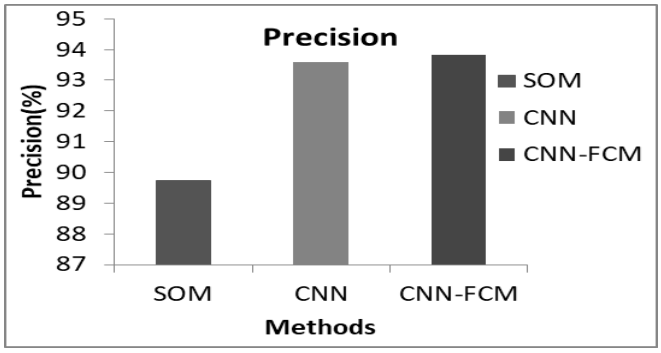

Figure 2. Precision outcomes of the suggested CNN-FCM approach and traditional approaches for KD prediction.

According to the suggested CNN-FCM, the precision comparison results depicted in Figure 2. The results shows that ABC features selection may use to forecast the KD accurately. The number of usable features in the proposed ABC does not significantly impact the performance of the jointly learned feature transformation. To address the categorization challenge, ABC employs a highly efficient approach given by ABC.

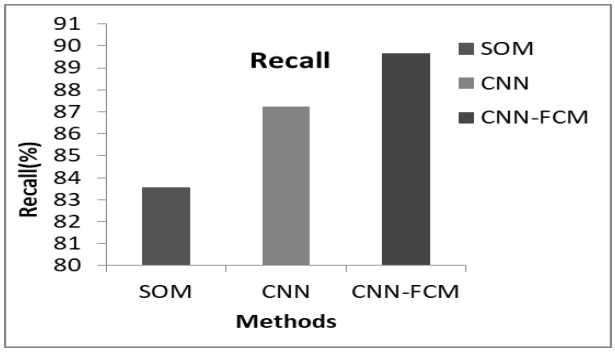

Figure 3. Recall comparison results of the proposed CNN-FCM approach and traditional approaches.

The suggested CNN-FCM method's accuracy outcomes are depicted in Figure.3. The findings show that the existing methods, such as CNN and SOM, have lower recall percentages than the one presented, at 89.68%, and the suggested approach has higher recall percentages of 90.78%.

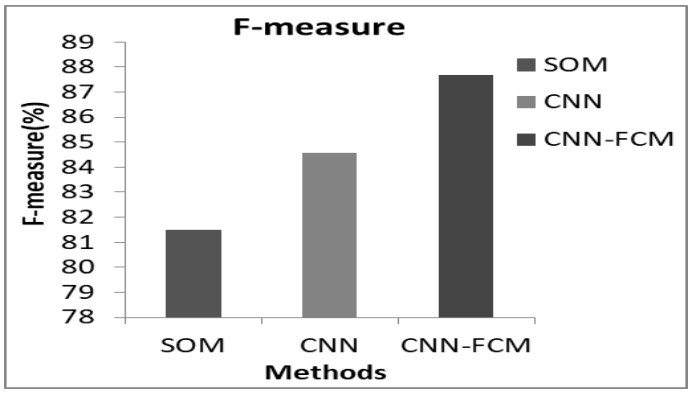

Figure 4. F-measure outcomes of the suggested CNN-FCM approach and traditional approaches for KD prediction

It can be seen in figure 4 that the suggested Fuzzy C-means clustering combined with the CNN classifier provides great quality in terms of the disease prediction rate, significantly superior to the CNN and SOM. A quantitative study employing the F-measure and machine learning methodologies yielded similar conclusions to the qualitative investigation. According to KD prediction accuracy, the proposed CNN-FCM is compared to traditional classification methods.

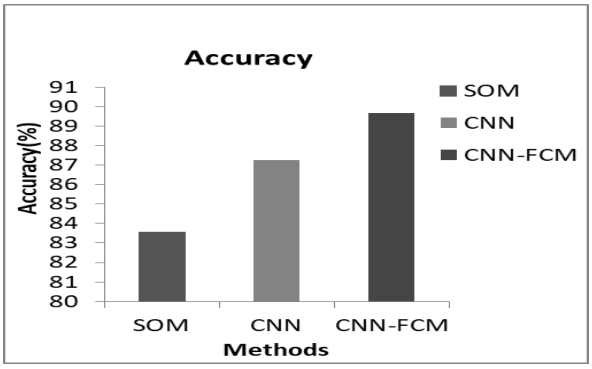

Figure .5. Accuracy comparison outcomes of the suggested CNN-FCM approach and traditional approaches for KD prediction

In Figure 5. the accuracy of the suggested CNN-FCM approach is compared with the traditional classifying approach. The findings show that the proposed approach is extremely efficient. Consequently, the classifiers will have a greater accuracy level than those

CONCLUSION

The proposed hybridization algorithm was used to analyze data to forecast IVIG resistance in KD patients. Initially, the RF procedure is presented for missing data restoration. Then to select the ideal feature from the training data set, the optimization process, the artificial bee colony approach. Finally, utilizing the proposed hybridization strategy, classification is performed after clustering and before classification. FCM clustering and Deep neural network classification are used here. The evaluation unit was used to adjust the hyperparameters. The CNN-FCM was proven to be the most efficient technique. In this study, we found out the importance of characteristics in predicting IVIG resistance. In KD patients, to predict IVIG resistance, this method uses deep learning. Consider incorporating this strategy into an EHR as clinical decision support shortly. Nonetheless, new data, particularly genetic variants, may assist us in enhancing our approach.

REFERENCES

[1] Makino N, Nakamura Y, Yashiro M, Kosami K, Matsubara Y, Ae R, et al. Nationwide epidemiologic survey of Kawasaki disease in Japan, 2015–2016. Pediatrics International. 2019;61(4):397-403.

[2] Kato H. Cardiovascular complications in Kawasaki disease: coronary artery lumen and long-term consequences. Progress in Pediatric cardiology. 2004 Dec 1;19(2):137-45. https://doi.org/10.1016/j.ppedcard.2004.08.007

[3] Benseler SM, McCrindle BW, Silverman ED, Tyrrell PN, Wong J, Yeung RS. Infections and Kawasaki disease: implications for coronary artery outcome. Pediatrics. 2005 Dec 1;116(6):e760-6. https://doi.org/10.1542/peds.2005-0559

[4] Honkanen VE, McCrindle BW, Laxer RM, Feldman BM, Schneider R, Silverman ED. Clinical relevance of the risk factors for coronary artery inflammation in Kawasaki disease. Pediatric cardiology. 2003 Mar;24(2):122-6. https://doi.org/10.1007/s00246-002-0063-1

[5] Cimaz R, Sundel R. Atypical and incomplete Kawasaki disease. Best Practice & Research Clinical Rheumatology. 2009 Oct 1;23(5):689-97. https://doi.org/10.1016/j.berh.2009.08.010

[6] Sonobe T, Kiyosawa N, Tsuchiya K, Aso S, Imada Y, Imai Y, Yashiro M, Nakamura Y, Yanagawa H. Prevalence of coronary artery abnormality in incomplete Kawasaki disease. Pediatrics International. 2007 Aug;49(4):421-6. https://doi.org/10.1111/j.1442-200X.2007.02396.x

[7] Dajani AS, Taubert KA, Gerber MA, Shulman ST, Ferrieri P, Freed M, Takahashi M, Bierman FZ, Karchmer AW, Wilson W. Diagnosis and therapy of Kawasaki disease in children. Circulation. 1993 May;87(5):1776-80. https://doi.org/10.1161/01.CIR.87.5.1776

[8] McCrindle BW, Li JS, Minich LL, Colan SD, Atz AM, Takahashi M, Vetter VL, Gersony WM, Mitchell PD, Newburger JW. Coronary artery involvement in children with Kawasaki disease: risk factors from analysis of serial normalized measurements. Circulation. 2007 Jul 10;116(2):174-9. https://doi.org/10.1161/CIRCULATIONAHA.107.690875

[9] Takeuchi M, Inuzuka R, Hayashi T, Shindo T, Hirata Y, Shimizu N, Inatomi J, Yokoyama Y, Namai Y, Oda Y, Takamizawa M. Novel risk assessment tool for immunoglobulin resistance in Kawasaki disease: application using a random forest classifier. The Pediatric infectious disease journal. 2017 Sep 1;36(9):821-6. https://doi.org/10.1097/INF.0000000000001621

[10] Zandstra J, Van de Geer A, Tanck MW, van Stijn-Bringas Dimitriades D, Aarts CE, Dietz SM, Van Bruggen R, Schweintzger NA, Zenz W, Emonts M, Zavadska D. Biomarkers for the discrimination of acute Kawasaki disease from infections in childhood. Frontiers in Pediatrics. 2020 Jul 22; 8:355. https://doi.org/10.3389/fped.2020.00355

[11] Bozlu G, Karpuz D, Hallioglu O, Unal S, Kuyucu N. Relationship between mean platelet volume-to-lymphocyte ratio and coronary artery abnormalities in Kawasaki disease. Cardiology in the Young. 2018 Jun;28(6):832-6. https://doi.org/10.1017/S1047951118000422

[12] Tremoulet AH, Dutkowski J, Sato Y, Kanegaye JT, Ling XB, Burns JC. Novel data-mining approach identifies biomarkers for diagnosis of Kawasaki disease. Pediatric research. 2015 Nov;78(5):547-53. https://doi.org/10.1038/pr.2015.137

[13] Sosa T, Brower L, Divanovic A. Diagnosis and management of Kawasaki disease. JAMA pediatrics. 2019 Mar 1;173(3):278-9. https://doi.org/10.1001/jamapediatrics.2018.3307

[14] Burns JC, DeHaan LL, Shimizu C, Bainto EV, Tremoulet AH, Cayan DR, Burney JA. Temporal clusters of Kawasaki disease cases share distinct phenotypes that suggest response to diverse triggers. The Journal of Pediatrics. 2021 Feb 1; 229:48-53. https://doi.org/10.1016/j.jpeds.2020.09.043

[15] Pezoulas VC, Papaloukas C, Veyssiere M, Goules A, Tzioufas AG, Soumelis V, Fotiadis DI. A computational workflow for the detection of candidate diagnostic biomarkers of Kawasaki disease using time-series gene expression data. Computational and Structural Biotechnology Journal. 2021 Jan 1; 19:3058-68. https://doi.org/10.1016/j.csbj.2021.05.036

[16] Kuniyoshi Y, Tokutake H, Takahashi N, Kamura A, Yasuda S, Tashiro M. Comparison of machine learning models for prediction of initial intravenous immunoglobulin resistance in children with Kawasaki disease. Frontiers in Pediatrics. 2020 Dec 3; 8:570834. https://doi.org/10.3389/fped.2020.570834

[17] Malhotra A, Joshi S. Exploring the Intersection of Demographic Change and Healthcare Utilization: An Examination of Age-Specific Healthcare Needs and Service Provision. Progression Journal of Human Demography and Anthropology. 2025 Jan 29:8-14.

[18] Ling XB, Lau K, Kanegaye JT, Pan Z, Peng S, Ji J, Liu G, Sato Y, Yu TT, Whitin JC, Schilling J. A diagnostic algorithm combining clinical and molecular data distinguishes Kawasaki disease from other febrile illnesses. BMC medicine. 2011 Dec 6;9(1):130. https://doi.org/10.1186/1741-7015-9-130

[19] Asha SJ, Pai P, Kamath P, Madli R, Arjunan RV. An Improved EEG Signal Feature Selection Paradigm for Migraine Detection. Journal of Internet Services and Information Security. 2024;14(3):143-56. https://orcid.org/0009-0000-4458-6806

[20] Chu FA, Xiang-qian HE, Yue YU, Jie TI, Sheng ZH, Zhe LI. BP neural network model for the differentiation of Kawasaki disease and febrile illnesses based on data mining. Chinese Journal of Evidence-Based Pediatrics. 2017 Feb 25;12(1):22.

[21] Pillai D, Bhatia S. Ontology-Driven Approaches for Standardizing Rare Disease Terminology. Global Journal of Medical Terminology Research and Informatics. 2024 Jun 28;2(2):5-9.

[22] Wang H, Tan X, Huang Z, Pan B, Tian J. Mining incomplete clinical data for the early assessment of Kawasaki disease based on feature clustering and convolutional neural networks. Artificial Intelligence in Medicine. 2020 May 1; 105:101859. https://doi.org/10.1016/j.artmed.2020.101859

[23] Shrivastav P, Malakar U. Exploring Barriers to Medication Adherence Among Patients with Chronic Diseases. Clinical Journal for Medicine, Health and Pharmacy. 2024 Sep 30;2(3):21-31.

[24] Palanisamy S, Kanmani S. Artificial bee colony approach for optimizing feature selection. International Journal of Computer Science Issues (IJCSI). 2012 May 1;9(3):432.

[25] Bhatia M, Ramesh Iyer R. Immunological Responses to Viral Infections. Medxplore: Frontiers in Medical Science. Periodic Series in Multidisciplinary Studies. 2025:52-70.

[26] Kim P. Convolutional neural network. InMATLAB deep learning: with machine learning, neural networks and artificial intelligence 2017 Jun 16 (pp. 121-147). Berkeley, CA: Apress. https://doi.org/10.1007/978-1-4842-2845-6_6

[27] Manlhiot C, Mueller B, O'Shea S, Majeed H, Bernknopf B, Labelle M, Westcott KV, Bai H, Chahal N, Birken CS, Yeung RS. Environmental epidemiology of Kawasaki disease: linking disease etiology, pathogenesis and global distribution. PloS one. 2018 Feb 7;13(2): e0191087. https://doi.org/10.1371/journal.pone.0191087

[28] Liu MY, Liu HM, Wu CH, Chang CH, Huang GJ, Chen CA, Chiu SN, Lu CW, Lin MT, Chang LY, Wang JK. Risk factors and implications of progressive coronary dilatation in children with Kawasaki disease. BMC pediatrics. 2017 Jun 6;17(1):139. https://doi.org/10.1186/s12887-017-0895-8

[29] Szymanski AM, Clifford H, Ronis T. Fever of unknown origin: a retrospective review of pediatric patients from an urban, tertiary care center in Washington, DC. World J Pediatr. 2020;16(2):177-184.

[30] Tang F, Ishwaran H. Random Forest missing data algorithms. Statistical Analysis and Data Mining: The ASA Data Science Journal. 2 https://doi.org/10.1002/sam.11348017 Dec;10(6):363-77. https://doi.org/10.1002/sam.11348

[31] Cannon RL, Dave JV, Bezdek JC. Efficient implementation of the fuzzy c-means clustering algorithms. IEEE transactions on pattern analysis and machine intelligence. 1986 Mar 31(2):248-55. https://doi.org/10.1109/TPAMI.1986.4767778

[32] Miriyala, G. P., & Sinha, A. K. (2025). Precision Diagnosis of Coronary Artery Disease with OTLGBM. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 16(1), 230-246. https://doi.org/10.58346/JOWUA.2025.I1.014