International Transactions on Electrical Engineering and Computer Science
N. Kumar et.al., Vol: 4, No: 1, pp: 38-52, March 2025

Optimizing Pre-Trained Models for Medical Dataset Classification with a Fine-Tuning Approach

N. Kumar [1,2], T. Christopher[3]

Abstract: Medical organizations struggle to deal with huge high-dimensional datasets that need powerful machine learning systems to produce precise healthcare outcomes. Traditional analytical techniques prove inadequate when dealing with extraction from features and performance of classifiers in this specific setting. The research introduces an algorithm which enhances Stacked Autoencoders (SAEs) by combining them with a customized Logistic Regression model intended for medical high-dimensional data analysis. This approach implements a Hybrid Imputation Method using MICE and KNN Imputation which precedes other stages and helps process missing values and outliers in medical data. We use CNNs and SAEs together for deep feature extraction before using Feature Fusion to assemble a robust feature collection. A set of the most important features is identified by executing Advanced Ensemble Feature Selection (EFS) procedures which include Few-shot Learning and Model-Agnostic Meta-Learning Algorithm (MAML) and Genetic Algorithm-Based Feature Selection (GAFS). The procedure of fine-tuning pre-trained models represents an effective enhancement for classification tasks particularly in situations of limited dataset availability. The experimental outcomes demonstrate remarkable performance gains in terms of accuracy and sensitivity and specificity as well as reduced execution time as compared to current techniques. Upcoming work for this study involves speeding up algorithm processing abilities and scalability alongside the integration of robust deep learning structures with self-supervised learning methodologies together with upgrade transfer learning approaches for medical dataset variety applications. The study will concentrate on enhancing model transparency through explainable AI and real-time validation for clinical deployment and ethical and regulatory compliance to develop this technique for practical healthcare settings.

Key words: Medical Dataset Classification, Real-Time Clinical Implementation, Convolutional Neural Network, Stacked Autoencoders, Logistic Regression classification model,

History
Received: 18-11-2024; Revised: 05-02-2025;
Accepted: 10-03-2025

N. Kumar
pkn.kumaar@gmail.com

[1]Department of Computer Science, Government Arts College, Udumalpet, Tamil Nadu – 642126, India
[2]Department of Computer Science, Dr. N. G. P Arts and Science College, Coimbatore – 641035, India
[3]Department of Computer Science, Government Arts College, Coimbatore- 641018, Tamil Nadu, India

1. Introduction

Medical data sets with class unbalance cause the creation of biased models for analysis purposes. The study presents SMOTE-RF-CS-LR as a hybrid model which unites SMOTE together with Random Forest and Cost-Sensitive Logistic Regression [1]. The research tested the method across seventeen datasets where it validated high metrics and analyzed rapid training periods indicating suitable application opportunities in medical practice.

FL stands as a solution which enables device operators to train models together with distributed border devices while maintaining privacy protection [2]. This research investigates the use of FL with Internet of Things (IoT) edge device computational

https://doi.org/10.62760/iteecs.4.1.2025.126

resources for sleep stage classification. The experiment analyzed FL performance while comparing random weight initializations to the pre-trained weights derived from MAML for assessment of initialization impact on FL results. According to the obtained results pre-trained weights result in improved model convergence alongside better classification accuracy.

EEG data possesses unique characteristics among users that prevents the creation of generalized AI models to classify EEG signals in Brain Computing Interface (BCI) applications [3]. The quick task-learning capability and minimal data requirement has made MAML establish itself as an operational method. An analysis of MAML's capability to enhance BCI systems with accurate new user classification from limited data points appears in this conference paper. The researchers implement MAML to boost EEG classification generalization and adaptation through Multitask Learning (MTL) by treating individual users as separate tasks.

The goal of data-free meta-learning is to extract value from pre-existing model knowledge while discarding actual training data. Traditional approaches face limitations because they focus solely on parameters and do not leverage data understanding and deal poorly with big networks while demanding uniform network structures. To resolve these issues, we propose PURER, a comprehensive framework that includes ePisode cUrriculum inveRsion (ECI) [4] for generating adaptive pseudo episodes and invErsion calibRation following inner loop (ICFIL) for enhancing meta testing. ECI enables dynamic, difficulty-adjusted pseudo episode creation, while ICFIL improves performance during meta testing. Our extensive experiments highlight PURER's superior effectiveness across various real-world scenarios.

Applications such as text categorization, Object Detection (OD), and Image Classification (IC) have observed notable progress in Machine Learning (ML) techniques. Nevertheless, the majority depend on optimizers that use derivatives, like Stochastic Gradient Descent (SGD) [5], which might impede performance by becoming stuck in local optima. By examining new areas of the Search Space (SS), this work analyzes the use of bio-inspired Meta-Heuristic (MH) techniques to refine pre-trained weights and increase their efficacy. Through experiments on two classification tasks (image and text) across four datasets, we demonstrate that nature-inspired algorithms can achieve superior results compared to traditional pre-trained architectures. Additionally, we analyse various architectures to identify the most critical weights for fine-tuning.

This research aims to develop a robust algorithm that surpasses current limitations by integrating advanced techniques at every stage of the data processing pipeline. The proposed approach enhances pre-processing, employs sophisticated FE and FS methods, and optimizes classification models to achieve superior diagnostic accuracy, thereby improving clinical decision-making support.

The format of the paper is as follows: The literature on FE, FS, pre-processing, and classification techniques in medical diagnostics is reviewed in Section 2. Section 3 outlines the methodology for optimizing pre-trained models for medical dataset classification using a fine-tuning approach. Section 4 summarizes the results and discusses how it might influence further study.

2. Related Work

Christopher et al. (2023) proposed the combination of Twin Bounded Support Vector Machine (TBSVM) and Adaptive Firefly Optimization Algorithm (AFOA) in their work titled [6] for medical data classification improvement. K-Means Clustering (KMC) initiates data processing to manage missing values and errors present in the datasets. AFOA performs feature selection through optimal fitness value generation and TBSVM implements data classification through structural risk minimization with a regularization term for maximizing the margin. This method leads to better classification performance as well as shorter training duration. Experimental findings demonstrate that AFOA-TBSVM achieves superior results compared to current techniques through its enhanced accuracy levels and better sensitivity along with specificity and reduced execution times.

Christopher et al (2023) designed Ensemble Feature Selection (EFS) with Back Propagation Neural Networks (BPNN) to handle the existing problems in their research [7]. K-Means Clustering (KMC) first processes data to manage missing entries before EFS determines the best suited features through objective function evaluation. The EFS feature selection system unites multiple approach methods including EEHO (Entropy Elephant Herding Optimization), AFOA (Adaptive Firefly Optimization Algorithm) and EBFO (Entropy Butterfly Optimization Algorithm) to achieve superior results. A trained multilayer feed-forward neural network operated by the BPNN algorithm conducts classification of medical information obtained from the dataset. The experimental results demonstrate that the proposed EFS-BPNN method achieves superior accuracy and sensitivity together with better specificity along with faster execution times in comparison to existing methods.

N. Kumar et al. (2023) introduced the HRFBP Neural Network (NN) approach through their work in [8]. The K-Means Clustering method works initially to prepare datasets by managing error rates and handling missing values. Subsequently significant features are extracted through Modified PCA (MPCA). The Ensemble Feature Selection algorithm (EFS) implements three diverse methods namely Entropy Elephant Herding Optimization (EEHO), AFOA, and Entropy Butterfly Optimization Algorithm (EBFO) to achieve enhanced performance during feature selection. After processing medical data, the HRFBP algorithm derives weight values that aid precision in class label forecasting. Scientific experimentation demonstrates that HRFBP provides higher accuracy performance while using fewer execution resources than preceding algorithms.

The authors Nahato et al (2023) from [9] created a new algorithm to enhance diabetes dataset classification through advanced high-dimensional medical data processing. The Hybrid Imputation Method uses an integration of MICE and KNN to treat (MV) Missing Values and outliers as its initial procedure. Initial deep feature extraction through CNNs and Autoencoders enables Feature Fusion to produce an extensive feature collection. The combination of GAFS and MOEA and Relief-Based Methods within an Advanced Ensemble Feature Selection technique detects the most crucial features. A Hybrid Model utilizing Stacking and Boosting and Bagging alongside Neural Network Enhancements equipped with attention mechanisms along with Transfer Learning (TL) achieves classification tasks.

A total of eight fine-tuning approaches for convolutional networks with ImageNet pre-training received development by Spolaôr et al. (2024) in [10] to minimize unwanted bias from limited data usage. Different approaches exist regarding network design as well as learning rate and number of blocks with active learning for these methods. The research team performed tests using two medical image collections containing 104 and 200 dermoscopic photos to prove Deep Learning (DL) works consistently for detecting malignant from non-malignant growths regardless of limited data availability. Another benefit of this approach is its universal application to different areas which was proven through successful results on a dataset composed of 746 CT images pertaining to COVID-19 diagnosis.

This paper presented by Chauhan et al. in [11] described a method for COVID-19 detection through CXR image classification using DL technology. A CNN classifier under TL can distinguish between pictures of COVID-19 infections and regular health pictures. The DenseNet model becomes more accurate when performing early prevention work. F1-score forms one set of evaluation metrics together with precision and accuracy and recall. Automated testing of multiple combinations involving optimizers and learning rate schedulers and loss functions leads to determining the best setup. The combination of Adamax optimizer and Cross Entropy loss function and StepLR scheduler demonstrated the maximum accuracy at 98.45% for normal-healthy CXR and 98.32% for COVID-19 images.

The detection system for Diabetic Retinopathy (DR) based on pre-trained CNN was established by Mohammadian et al (2017) in [12]. Detouring the lengthy process of training a CNN from beginning makes pre-trained networks the chosen option. Each network performs fine-tuning on the same pre-processed dataset and the researchers evaluate both fine-tuning parameters and pre-trained network choices. The study provides an expedited approach to optimize pre-trained networks for diabetes retinopathy testing which allows users to examine different adjustment parameters alongside their impact on operational performance.

The research paper authored by Roshan et al (2020) in [13] proposed performing two case studies with distinct data sets. The first research study decreased time consumption together with financial expenses by applying pre-trained convolutional neural networks instead of complete automatic learning on the Kaggle dataset. The networks received processed data for adjustment to achieve maximum performance in the selected parameters. The proposed method achieved resilience when the researchers applied its tuned network to diagnose 101 clinical photos that varied across different devices and picture quality and scale.

The research of Villa-Pulgarin et al. (2022) presented skin lesion classification through four experimental evaluations of CNN in [14]. The research analyzes the steps including (DA) Data Augmentation and fine-tuning optimization, which use TL with ImageNet weights for optimal results. The authors utilize HAM10000 data to develop three CNN solutions for comparison: DenseNet-201, Inception-ResNet-V2 and Inception-V3. The proposed models have reached prediction accuracy levels of 98%, 97% and 96%. A 93% success rate was obtained during testing the most effective model against the data from the ISIC 2019 dataset. Based on the results, this CNN-based technology demonstrates potential as a precise instrument for detecting skin cancer.

Liu et al. (2020) developed a TL framework which organized gastric M-NBI images into three diagnostic categories including early gastric cancer (EGC) and low-grade neoplasia (LGN) and chronic gastritis (CGT) as per [15]. The CNNs were fine-tuned pre-trained. Researchers assess this method through direct training of CNNs on their dataset combined with study of conventional handcrafted texture FE techniques. The experiment showed that using CNN with fine-tuning produced superior performance compared to direct training of CNNs alongside the use of traditional features. The ResNet50 model achieved respectively 0.96 accuracy and F1-scores 0.92 and 0.91 and 0.99 for CGT and LGN and EGC gastrointestinal classifications. The proposed framework delivers good results when analyzing multiple classes of gastric M-NBI IC.

The usage of linear SVM for classification was examined by Bal-Ghaoui et al (2023) through different fine-tuning methods on five ImageNet pre-trained state-of-the-art models comprising VGG16, VGG19, ResNet50, ResNet101, and InceptionV3 in [16]. The researchers utilized four publicly available ultrasound dataset collections to perform their experiments which included information about thyroid nodules and breast cancer as well as salivary gland cancer. Accuracy and precision together with recall established the performance evaluation metrics through five-fold stratified (CV) Cross-Validation methods. Performance optimization is possible through fine-tuning 15% of the final layers in both ResNet50 and InceptionV3 networks. Overall results benefit from using SVM as classifier for the best models by 6% thereby showing how choosing the right classifier method together with proper tuning leads to superior ultrasound IC performance.

3. Proposed Methodology

This section proposes the MAML approach to improve model training performance with small amounts of data. With the least amount of data possible, models may be quickly adapted to different tasks due to the meta-learning technique known as MAML. The key processes involved in the MAML approach are the training of model parameters through gradient descent, the simulation of few-shot learning scenarios with various tasks, and optimization to ensure effective fine-tuning. The overall system is depicted in Fig. 1.

3.1 *Input dataset collection*

The UCI ML Repository provided the datasets used in this study [17]. The Fertility dataset is among them. The Fertility dataset, sourced from 100 volunteers, includes semen samples analysed according to WHO 2010 criteria. Sperm concentration is correlated with environmental factors, lifestyle choices, health status, and sociodemographic information in this dataset, which is available in the UC Irvine ML Repository. Season, age, childhood illnesses, accidents or major trauma, surgical interventions, high fevers during the previous year, frequency of alcohol use, smoking behaviors, number of daily sitting hours, and diagnosis are among the 100 cases and 10 features that make up this dataset.

3.2 *Pre-processing using K-Nearest Neighbours (KNN) algorithm*

This work uses pre-processing with the k-nearest-neighbors (KNN) method to improve the accuracy of disease detection for the given database [18]. Fertility data is one of the unstructured datasets that are used. One of the easiest and simplest Data Mining (DM) methods is KNN. Because the training examples must be kept in memory during run-time, it is known as memory-based classification [19]. The Euclidean Distance (ED) is used to compute the difference between features when handling continuous attributes. The formula for calculating the distance between two instances, where the first is (a1, a2, a3, ... an) and the second is (b1, b2, b3, ... bn), is as follows (1)

$$d(i,j) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (1)$$

Here x_i and y_i in Euclidean n-space, are two points.

A significant issue when using the ED formula is that attributes with larger value ranges can dominate those with smaller ranges. For instance, age ranges between 40 and 80 and cholesterol levels between 100 and 190 may be observed in heart disease records. This implies that the distance calculation would be more impacted by cholesterol levels than by age. Continuous features are normalized to equalize their impact on the distance measure among instances in order to solve this issue. KNN can handle discrete attributes as well, even though it works better with continuous attributes. When working with discrete attributes, the difference is set to one if the attribute values of the two instances (a2 and b2) differ, and set to zero otherwise. In Fig. 2 and Fig. 3, The K-NN algorithm makes it simple to determine the category or class of any given dataset.

Algorithm 1: KNN algorithm

- Choose the neighbor with number k first.
- Based on the K number of neighbors, the ED must be determined.
- Determine how many data points in each category among these K neighbors.
- The category with the maximum number of neighbors will receive new data points.
- The K-NN classification model will then be prepared in this manner.

Equation (2) is the formula for calculating the ED, the Fig. 4, illustrates the ED among A_1 and B_1 =

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad (2)$$

In the pre-processing phase using the K-Nearest Neighbours (KNN) algorithm [20], the database is first divided into two sets: one with complete instances and one with incomplete instances that have missing values. For each incomplete instance, KNN identifies 'k' nearest complete instances based on similarity in the feature space. The mean or mode of the nearest values of these nearest neighbors is then used to impute the MV. After imputation, the instance is validated to ensure it fits correctly within its cluster. If not, alternative values are tested iteratively until the correct cluster is achieved. This method enhances the overall data quality and classification accuracy by ensuring that each instance is accurately represented, making it particularly effective for medical datasets where precise classification is crucial.

3.3 Feature Extraction using Convolutional Neural Networks (CNN) and Stacked Autoencoders (SAEs) algorithm

FE using CNN and Stacked Autoencoders (SAEs) enhances data analysis by leveraging deep learning. CNNs capture spatial hierarchies and invariant features from images using convolutional, pooling, and activation layers, maintaining computational efficiency. SAEs learn efficient data coding's in an unsupervised manner, with each layer progressively abstracting data through encoding and decoding. Integrating CNNs as encoders in SAEs captures both low-level and high-level features, improving data representation and dimensionality reduction. This combination excels in image recognition, noise reduction, and anomaly detection, enhancing accuracy and feature differentiation and paving the way for further advancements in deep learning architectures.

3.3.1 Convolutional Neural Networks (CNN)

Because of their strong FE capabilities, CNN are commonly utilized in image-based learning applications. Convolutional Layers (CL), which use filters to identify features like edges and textures, and pooling layers, which reduce the spatial dimensions of Feature Maps (FM) to lessen computational load and avoid overfitting, are commonly seen in CNN [21] designs. Max pooling, which chooses the highest value from each patch, and average pooling, which calculates the average, are the two primary forms of pooling. High-level features are used by Fully Connected (FC) layers for classification tasks. CNNs demonstrate high accuracy and efficiency in image recognition, noise reduction, and anomaly detection. The CL, pooling layer, and FC layer are the three basic layers that make up the CNN architecture.

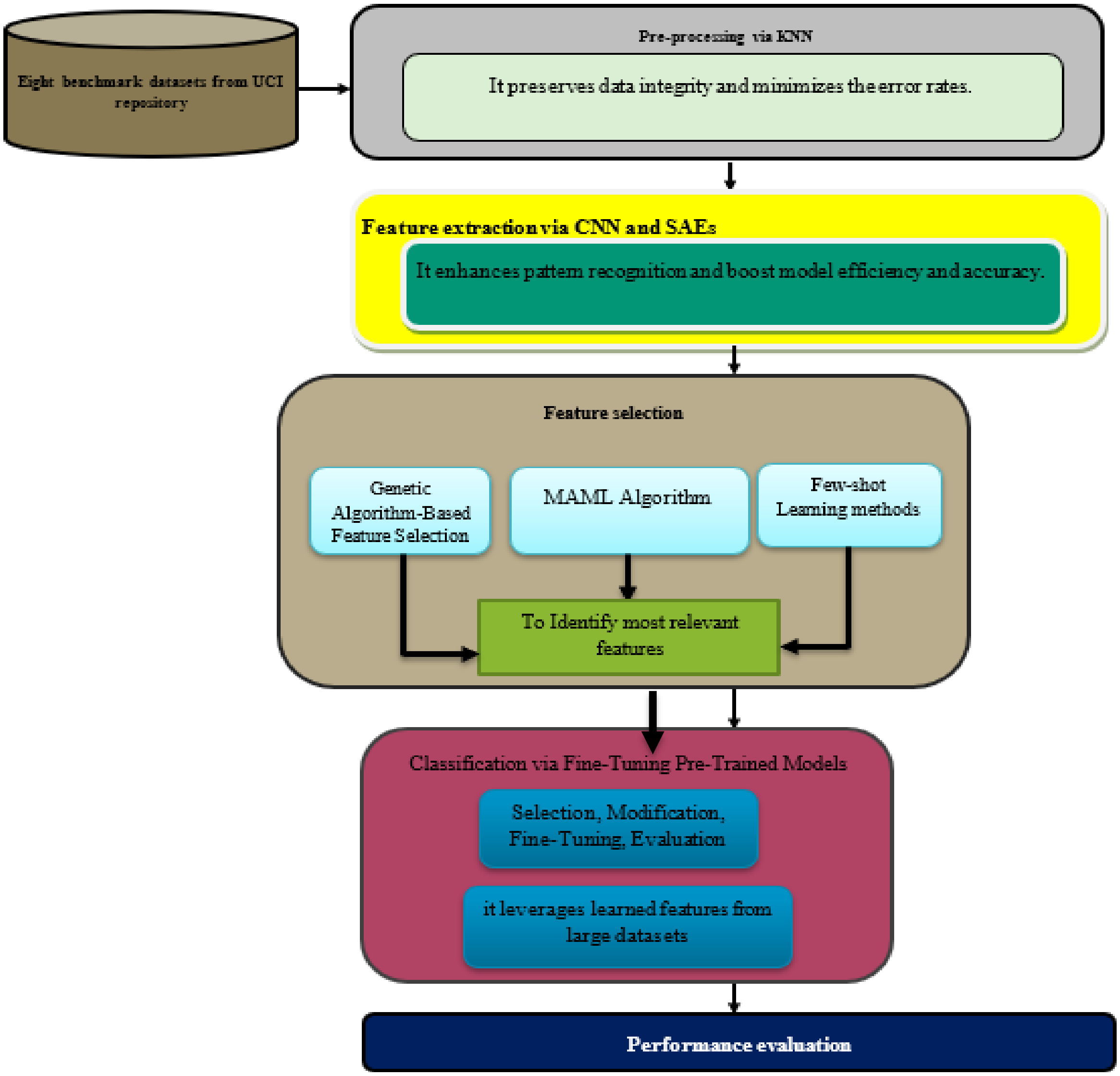

Fig .1: SAEs – MAML algorithm's overall block diagram

Applying a filter that slides over the input image to create a FM is the process by which convolutional layers [22] extract meaningful visual features from an image. Equation (3) demonstrates the manner in which this process is represented mathematically by the convolution operation:

$$F = I * K \tag{3}$$

where, F is the output feature map, I is the input image, and K is the filter. The symbol $*$ denotes the convolution operation, which involves calculating the dot product between the filter and the corresponding section of the input image. The Fig. 5 illustrates the Feature Extraction and classification.

3.3.2 *Stacked Autoencoders (SAEs)*

Stacked Autoencoders (SAEs) [23] are an advanced form of autoencoders, designed for nonlinear dimensionality reduction and feature learning. They consist of multiple autoencoders stacked together, where each layer is trained to compress and reconstruct data progressively. The first autoencoder compresses input data into a latent space, which serves as input for the next autoencoder, and so on. After layer-wise training, the entire stack is fine-tuned. SAEs capture complex, nonlinear relationships and hierarchically learn features, making them effective for applications like noise reduction, anomaly detection, and data compression

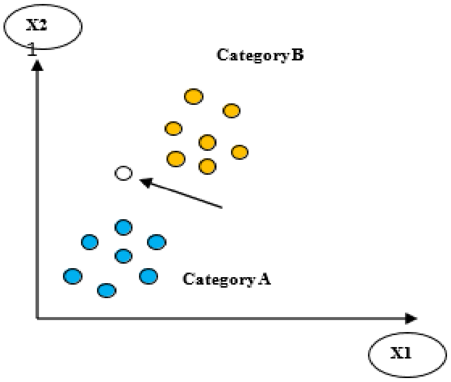

Fig. 2: Before KNN

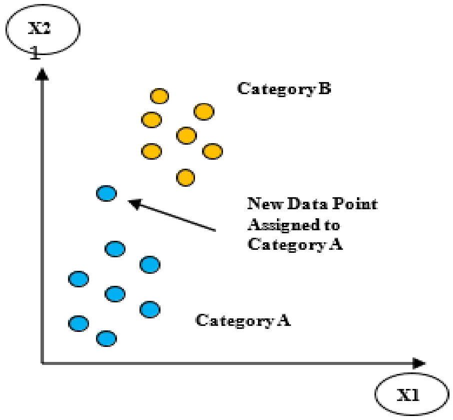

Fig. 3: After KNN

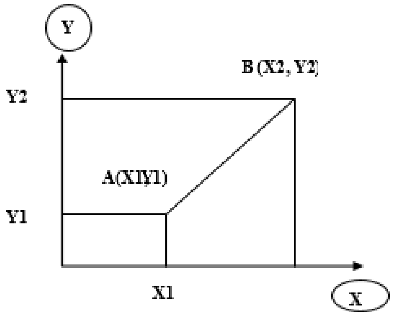

Fig. 4: Euclidian distance between A1 and B1

The Fig. 6 illustrates the Stacked Autoencoders (SAEs) [24] should start with an "Input Data" layer on the left, represented as nodes. An encoder condenses the input into a smaller "Hidden Representation" layer, and a decoder uses this representation to recreate the original input. This is how an autoencoder works. Connect encoders and decoders vertically, showing how each autoencoder's output serves as the next autoencoder's input. The final decoder produces the "Reconstructed Data" or output. Use arrows to indicate data flow and optionally label activation functions like ReLU or sigmoid for clarity.

3.4 Feature Selection using GAFS, MAML and Few-shot Learning Methods

The Feature Selection method leverages three key techniques to enhance feature relevance identification. Genetic Algorithm-Based Feature Selection (GAFS) applies evolutionary principles to iteratively select optimal features through simulated natural selection. Model-Agnostic Meta-Learning Algorithm (MAML) optimizes multiple criteria simultaneously, ensuring a balanced and comprehensive feature set. Few-shot Learning methods assess feature relevance based on their ability to differentiate between similar instances with minimal examples. By integrating these approaches, the method aims to outperform individual techniques, offering a more robust and effective feature selection process.

3.4.1. Genetic Algorithm Based Feature Selection (GAFS)

The genetic algorithm [25] is a cutting-edge feature selection method rooted in principles of natural genetics and evolution. As depicted in Fig. 7, this algorithm employs several key operators: initialization, fitness assignment, selection, crossover, and mutation.

Initialization**:** The process starts by creating and randomly setting the genes of individuals in the population, as the algorithm relies on stochastic methods for optimization.

Fitness Assignment: After initializing the population, each individual's fitness is evaluated by training neural networks with training data and assessing performance with testing data. Poor performance results in lower fitness. We use a rank-based method to assign fitness scores, favouring individuals with higher scores for recombination.

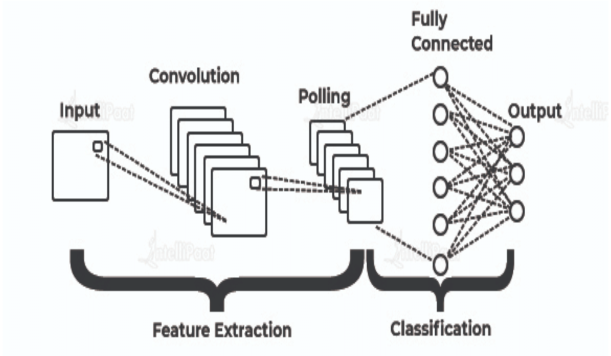

Fig. 5: CNN Architecture

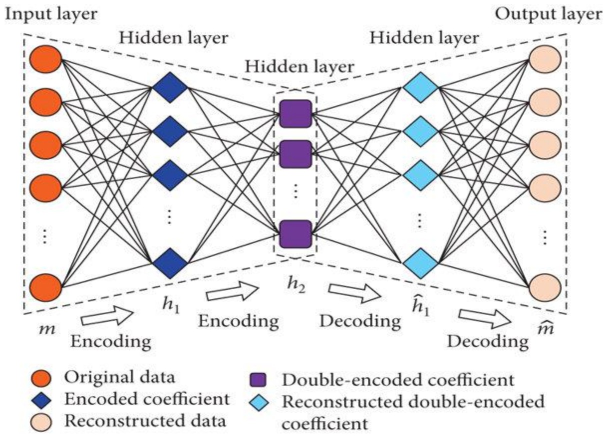

Fig. 6: Stacked Autoencoders (SAEs).

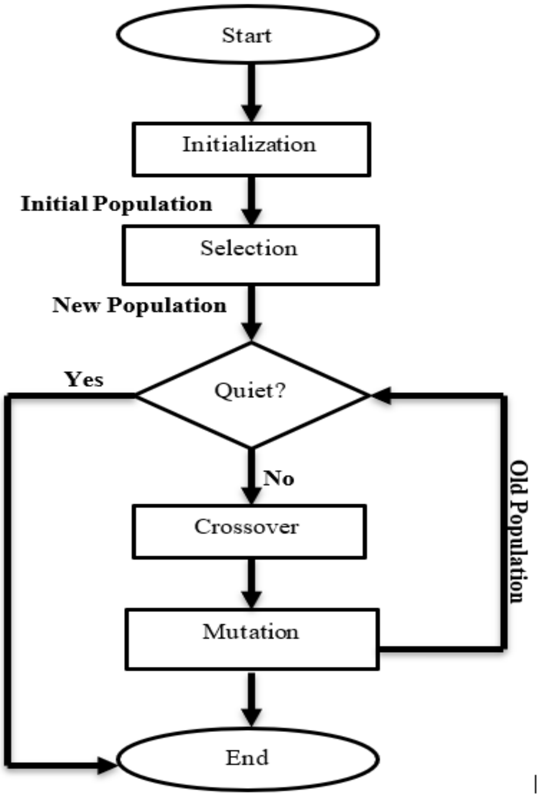

Fig. 7: GA feature selection

Selection: Individuals are selected for the next generation based on their fitness levels. The selection process, using stochastic sampling, picks individuals proportionally to their fitness. The number of selected individuals is half the population size (N/2).

Crossover: To generate a new population, the crossover operator combines the genes of two randomly chosen individuals. The uniform crossover method decides whether each trait of the offspring is inherited from one or both parents.

Mutation: To introduce variability and prevent similarity among offspring, the mutation operator randomly alters some features. This is achieved by generating a random number between 0 and 1 to determine if a feature should be mutated.

3.4.2. Model-Agnostic Meta-Learning Algorithm (MAML)

Models can quickly adapt to new tasks with minimum input due to a meta-learning framework called the MAML [26], created by Chelsea Finn et al. in 2017. The objective of MAML is to identify the ideal set of model parameters that are effectively adjustable for a range of applications. There are two primary steps to this approach:

- **Meta-Training:** After a few gradient updates, the model is trained on various tasks to maximize its performance. This process involves computing the gradients for each task and updating the model parameters to minimize the aggregate loss across all tasks.
- **Task-Specific Adaptation:** For each new task, the model performs a few gradient descent steps using task-specific data to adapt its parameters. This is known as the inner loop of optimization.
- **Meta-Optimization:** The model's initial parameters are updated based on the performance of the adapted model on all tasks, aiming to improve generalization across tasks [27]. This is known as the outer loop of optimization.

3.4.3. Few – shot Learning Methods

Few-shot learning methods [28] allow frameworks to perform well with very limited data. Metric-based methods, like Prototypical Networks, Siamese Networks, and Matching Networks, rely on similarity functions to classify new instances based on their proximity to known examples. Meta-learning approaches, such as MAML, Reptile, and Meta-SGD, focus on training models to rapidly adapt to new tasks using minimal examples as illustrated Fig. 8.

Optimization-based methods, including Transfer Learning and Fine-Tuning, adjust pre-trained models for specific tasks with few-shot data [29]. Data augmentation methods create synthetic examples or modify existing ones to enhance the training set, while generative methods like GANs and VAEs generate new data to improve learning. Attention mechanisms, through techniques like Memory-Augmented Networks and Transformers, highlight relevant features to boost performance in few-shot scenarios. Each method offers distinct advantages depending on the task requirements and the amount of available data.

3.5 Fine Tuning Pre-Trained Classification Model

Fine-tuning a pre-trained classification model [30] is a refined transfer learning approach designed to enhance model performance on specific tasks. This method involves adapting a base network, initially trained on a large, diverse dataset to learn general features, to a new, often smaller dataset. The base network's layers are transferred to a target network, with the first few layers typically frozen to preserve the learned features. The remaining layers are either randomly initialized or further trained to address the specific task at hand. Fine-tuning strategies vary, including updating the pre-trained layers through backpropagation of new task errors or selectively freezing layers based on their relevance to the new task. This approach leverages the pre-existing general features captured by the base network while customizing the model for the particular nuances of the new dataset. Recent advancements also explore dynamic strategies for layer selection and adaptation, integrating optimization techniques to refine which layers are fine-tuned, thereby improving the model's accuracy and robustness for specialized classification tasks.

3.5.1. Transfer Learning

A Machine Learning (ML) technique called TL [31] involves adapting a model created for one task to perform well on another that is related to it. When the target domain has little labelled data, it makes use of information from the source domain to enhance learning in the target domain.

3.5.2 Key Concepts of Transfer Learning

By transferring information from a source domain to a target domain, TL [32] makes use of pre-trained models to perform new tasks with little data. Pre-training and fine-tuning are usually the two main phases of this method. In order to acquire broad features, a model is pre-trained using a sizable diverse dataset. A base for TL is provided by this pre-trained model.

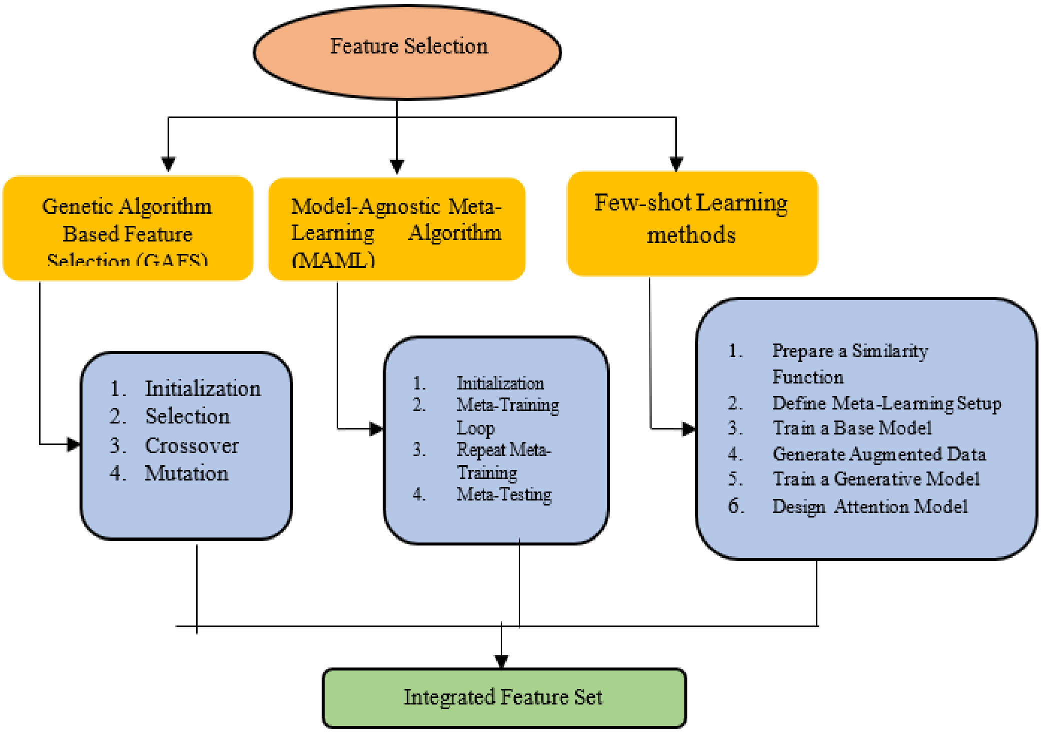

Fig. 8: Feature selection flow diagram

The model is adjusted to a particular target task during the fine-tuning phase. This entails updating all or just a portion of the model's layers to fit the new task's data, which is often more compact and domain-specific. FE is a popular technique that involves using the pre-trained model's layers to extract features, then training a new classifier on top of these features without changing the pre-trained layers. Additionally, domain adaptation techniques are employed to adjust models to handle discrepancies between the source and target domains, which may include modifying the model architecture or learning mechanisms to better align the domains. Transfer learning strategies can therefore be broadly categorized into fine-tuning, feature extraction, and domain adaptation, each serving different purposes in enhancing model performance on new tasks.as illustrated the Fig. 9.

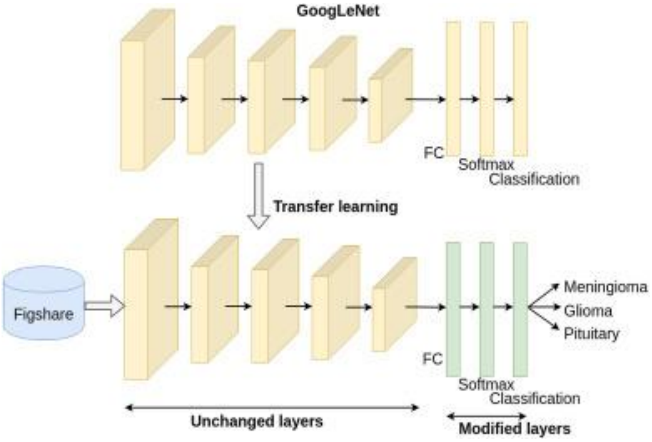

Fig. 9: Transfer learning using deep CNN

This comprehensive approach to fine-tuning pre-trained models [33], incorporating various techniques and strategies, significantly improves the robustness and accuracy of classification tasks, particularly in medical data analysis. To assess the efficiency of fine-tuned models, several metrics are used, including accuracy, sensitivity (recall), specificity, precision, F-measure, and execution time. For instance, in a study using four datasets from the UCI ML repository [34], the models were pre-processed using Fine Tuning Pre-Trained Based (MICE and KNN), feature extraction combining CNN and Stack Autoencoders, and feature selection employing GAFS, MAML, and Few – Shot Learning methods. Using these parameters, the performance of the Fine-tuned models was evaluated, showing that fine-tuning can improve classification efficiency, accuracy, sensitivity, specificity, and execution time when compared to current and suggested techniques. Table. 1 presents the outcomes of the performance comparisons.

Accuracy

The total actual classification parameters $(T_p + T_n)$, which are divided by the sum of the classification parameters $(T_p + T_n + F_p + F_n)$, constitute equation (4), which is used to calculate accuracy, which is defined as the overall correctness of the model. Table. 1 is used to compute the accuracy.

$$Accuracy = \frac{T_p+T_n}{(T_p+T_n+F_p+F_n)} \quad (4)$$

Where

Fn stands for False negative, Fp stands for False positive

Tp stands for True positive, and Tn stands for True negative

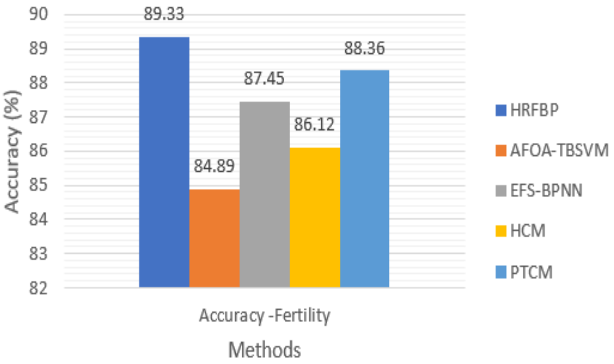

Fig. 10: Accuracy

As described in Fig. 10, the performance analysis graph compares the accuracy of five classification models: HRFBP, AFOA-TBSVM, EFS-BPNN, HCM, and PTCM. The HRFBP model demonstrates the highest accuracy at 89.33%, highlighting its strong classification performance. The PTCM follows with an accuracy of 88.36%, indicating its effectiveness in handling the dataset. The EFS-BPNN achieves an accuracy of 87.45%, while the HCM model records an accuracy of 86.12%, showing consistent performance. The AFOA-TBSVM model, with an accuracy of 84.89%, reflects relatively lower performance compared to the other models. Overall, the HRFBP model's superior accuracy underscores its robustness and reliability in classification tasks.

Table. 1: Results of performance comparison

Metrics and dataset	Methods				
	HRFBP	**AFOA-TBSVM**	**EFS-BPNN**	**HCM**	**PTCM**
Accuracy - Fertility	89.33	84.89	87.45	86.12	88.36
Recall - Fertility	88.42	85.11	86.23	87.56	87.90
Precision - Fertility	87.29	83.20	86.45	85.72	86.85
F-measure -- Fertility	86.79	84.14	85.14	85.90	86.15
Execution time - Fertility	15.19	19.45	17.62	16.83	18.50

Recall

Recall or sensitivity, often referred to as the TP rate, recall, or probability of detection in different measurement domains, is the proportion of T_P that are correctly identified as true as shown in equation (5).

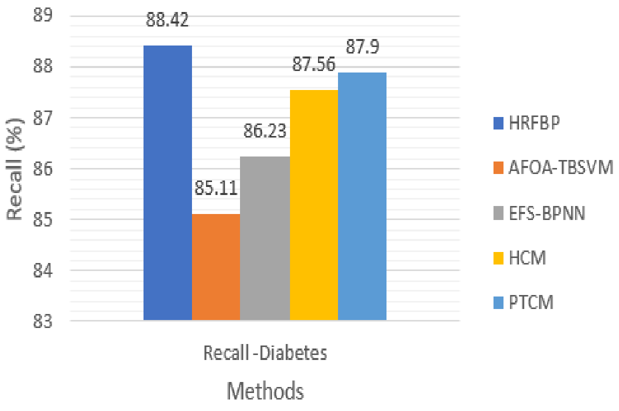

Fig. 11: Recall

$$Recall \;= \frac{T_p}{T_p+F_n} \qquad (5)$$

The performance analysis graph illustrates the Fig. 11, the recall percentages across five classification models: HRFBP, AFOA-TBSVM, EFS-BPNN, HCM, and PTCM. The HRFBP model achieves the highest recall at 88.42%, highlighting its strong ability to correctly identify positive cases. The PTCM follows with a recall of 87.90%, reflecting its consistent performance. The HCM model records a recall of 87.56%, indicating reliable classification results. The EFS-BPNN achieves a recall of 86.23%, while the AFOA-TBSVM shows the lowest recall at 85.11%. Overall, the HRFBP model's highest recall underscores its robustness and effectiveness in accurately identifying cases.

Precision

Precision refers to the percentage of the results that are relevant and defined as equation (6),

$$Precision=\frac{Tp}{tp+fp} \qquad (6)$$

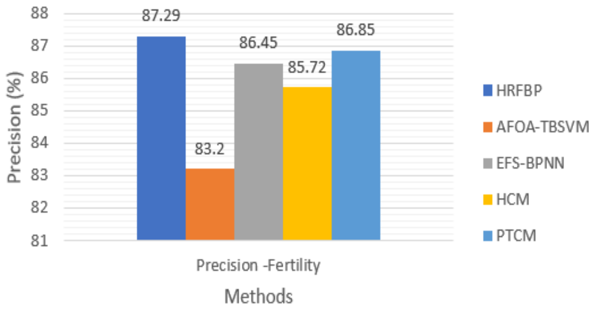

Fig. 12: Precision

Described in Fig. 12, the performance analysis graph presents the precision percentages across five classification methods: HRFBP, AFOA-TBSVM, EFS-BPNN, HCM, and PTCM. The HRFBP model demonstrates the highest precision at 87.29%, indicating its strong ability to correctly identify true positive cases among its predicted positive instances. The PTCM follows with a precision of 86.85%, reflecting reliable performance in identifying true positives. The EFS-BPNN achieves a precision of 86.45%, indicating a reasonable capability in predicting positive cases accurately. The HCM records a precision of 85.72%, showing slightly lower accuracy in identifying true positive cases. The AFOA-TBSVM exhibits the lowest precision at 83.20%, suggesting a higher likelihood of misclassification compared to other models. Overall, the HRFBP's highest precision highlights its superior ability to correctly identify true positive cases.

F – measure

The F-measure analysis graph for classification performance shows in Fig. 13, the results of five classification methods: HRFBP, AFOA-TBSVM, EFS-BPNN, HCM, and PTCM. The HRFBP model achieves the highest F-measure at 86.79%, demonstrating strong balanced predictive performance between precision and recall. The PTCM follows with an F-measure of 86.15%, reflecting reliable and consistent predictive capability. The HCM records an F-measure of 85.90%, showing slightly lower but still effective performance. The EFS-BPNN achieves an F-measure of 85.14%, indicating a reasonable ability to maintain a balance between precision and recall. The AFOA-TBSVM exhibits the lowest F-measure at 84.14%, suggesting room for improvement in overall predictive balance. Overall, the HRFBP's highest F-measure highlights its enhanced capability in accurately identifying true positive cases while minimizing false positives and false negatives.

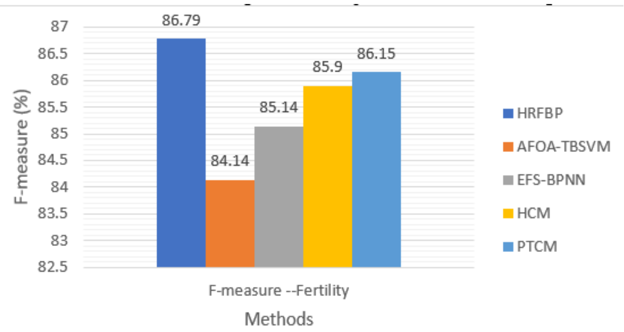

Fig. 13: F-measure

Execution Time

The system performs better when the recommended algorithm executes faster. Illustrated in Fig. 14, the graph compares the execution times of five methods: HRFBP, AFOA-TBSVM, EFS-BPNN, HCM, and PTCM for the dataset. The AFOA-TBSVM stands out with the highest execution time at 19.45%, indicating it is the least efficient among the methods. The EFS-BPNN follows with an execution time of 17.62%, reflecting moderate processing efficiency. The PTCM records an execution time of 18.50%, showing improved performance but still higher than some other methods. The HCM demonstrates an execution time of 16.83%, indicating relatively better efficiency. The HRFBP exhibits the lowest execution time at 15.19%, making it the most efficient among the compared models. Overall, the HRFBP's lowest execution time underscores its faster processing capability compared to the other methods.

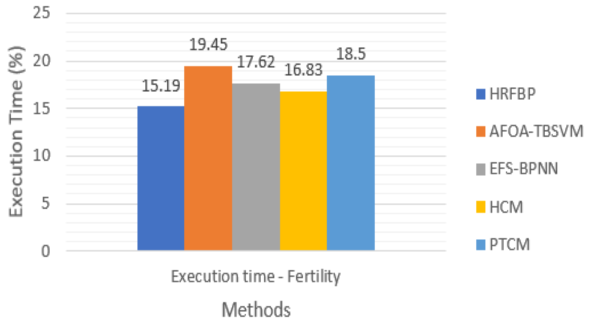

Fig. 14: Execution time

4. Conclusion

In this conclusion, this paper developed a novel algorithm that integrates Stacked Autoencoders (SAEs) with a logistic regression model to address the complexities of high-dimensional medical data. Our approach initiates with a Hybrid Imputation Technique combining MICE and KNN Imputation to effectively handle missing values and outliers. Deep FE utilizing CNN and SAE is then performed, bolstered by a strong Feature Fusion procedure. Then, in order to find the most pertinent features for better classification performance, we employ sophisticated EFS techniques including Few-shot Learning, MAML, and GAFS. The classification is further refined by Fine-Tuning Pre-Trained Models, which proved highly effective in scenarios with limited data. Our experiments show significant improvements in accuracy, sensitivity, specificity, and execution time compared to existing methods. Future work will focus on optimizing computational efficiency, integrating advanced deep learning and self-supervised learning techniques, refining transfer learning strategies, enhancing model transparency through explainable AI, and addressing real-time clinical use, ethical, and regulatory considerations for practical healthcare applications.

Conflict of interest

The authors declared "No conflict of interest"

Data set

https://www.kaggle.com/datasets/gabbygab/fertility-data-set

References

[1] A. Gupta, and S. Gupta "Enhanced Classification of Imbalanced Medical Datasets using Hybrid Data-Level, Cost-Sensitive and Ensemble Methods", *International Research Journal of Multidisciplinary Technovation*, Vol.6, No.3 pp. 58-76, 2024. https://doi.org/10.54392/irjmt2435

[2] S. Moon, T. S. Kim, J. Ryu and W. H. Lee, "Federated Learning for Sleep Stage Classification on Edge Devices via a Model-Agnostic Meta-Learning-Based Pre-Trained Model", *2023 IEEE 13th International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*, Berlin, Germany, pp. 188-192, 2023. https://doi.org/10.1109/ICCE-Berlin58801.2023.10375664

[3] M. H. Javaid, I. A. Shah, M. S. Javaid, U. Bin Irshad and Z. Halim "Model Agnostic Meta Learning for EEG Classification: Multitask Approach", *2023 IEEE IAS Global Conference on Emerging Technologies (GlobConET)*, London, United Kingdom, pp. 1-4, 2023. https://doi.org/10.1109/GlobConET56651.2023.10150186

[4] Z. Hu, L. Shen, Z. Wang, T. Liu, C. Yuan and D. Tao "Architecture, Dataset and Model-Scale Agnostic Data-free Meta-Learning", *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, pp. 7736-7745, 2023. https://doi.org/10.1109/CVPR52729.2023.00747

[5] G. H. De Rosa, M. Roder, J. P. Papa and C. F. G. Dos Santos "Improving Pre- Trained Weights through Meta - Heuristics Fine- Tuning", *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, Orlando, FL, USA, pp. 1-8, 2021. https://doi.org/10.1109/SSCI50451.2021.9659945

[6] T. Christopher, N. Kumar "Optimization Based Feature Selection Algorithm with Twin-Bounded Support Vector Machine for Medical Dataset Classification", *Journal of Survey in Fisheries Sciences*, Vol. 10, No. 4S, pp.1079-96, 2023. [CrossRef]

[7] T. Christopher, N. Kumar. Medical Dataset Classification Using Ensemble Feature Selection and Back Propagation Neural Network Algorithm", *International Journal of Intelligent Systems and Applications in Engineering*, Vol. 12, No. 22S, pp. 1403-1420, 2024. [CrossRef]

[8] P. Kumar et al. "Feature subset selection using filter, heuristic and meta-heuristic approaches using binary encoded diabetes dataset." *AIP Conference Proceedings*. Vol. 2555. No. 1, 2022. https://doi.org/10.1063/5.0108858

[9] K. B. Nahato, KH. Nehemiah, and A. Kannan "Hybrid approach using fuzzy sets and extreme learning machine for classifying clinical datasets", *Informatics in Medicine Unlocked*, Vol. 2 pp.1-11, 2016. https://doi.org/10.1016/j.imu.2016.01.001

[10] N. Spolaôr, H. D. Lee, A. I. Mendes, C. V. Nogueira, A. R. S. Parmezan, W. S. R. Takaki, C. S. R. Coy, F. C. Wu, and R. F. Pinto "Fine-tuning pre-trained neural networks for medical image classification in small clinical datasets", *Multimedia Tools and Applications*, Vol. 83, No. 9, pp. 27305-27329, 2024. https://doi.org/10.1007/s11042-023-16529-w

[11] T. Chauhan, H. Palivela, S. Tiwari "Optimization and fine-tuning of DenseNet model for classification of COVID-19 cases in medical imaging", *International Journal of Information Management and Data Insights*, Vol. 1, No. 2, art. no. 100020, 2021. https://doi.org/10.1016/j.jjimei.2021.100020

[12] S. Mohammadian, A. Karsaz and Y. M. Roshan, "Comparative Study of Fine-Tuning of Pre-Trained Convolutional Neural Networks for Diabetic Retinopathy Screening", *2017 24th National and 2nd International Iranian Conference on Biomedical Engineering (ICBME)*, Tehran, pp. 1-6, 2017. https://doi.org/10.1109/ICBME.2017.8430269

[13] SM. Roshan, A. Karsaz, AH. Vejdani, YM. Roshan "Fine-tuning of pre-trained convolutional neural networks for diabetic retinopathy screening: a clinical study", *International Journal of Computational Science and Engineering*, Vol. 21, No. 4, 564-573, 2020. https://doi.org/10.1504/IJCSE.2020.106869

[14] JP. Villa-Pulgarin, AA. Ruales-Torres, D. Arias-Garzon, MA. Bravo-Ortiz, HB. Arteaga-Arteaga, A. Mora-Rubio, et al "Optimized convolutional neural network models for skin lesion classification", *Computers Materials and Continua*, Vol. 70, No. 2, pp. 2131-2148, 2022.

https://doi.org/10.32604/cmc.2022.019529

[15] X. Liu, C. Wang, J. Bai, G. Liao “Fine-tuning pre-trained convolutional neural networks for gastric precancerous disease classification on magnification narrow band imaging images”, *Neurocomputing*, Vol. 392, pp. 253-267, 2020. https://doi.org/10.1016/j.neucom.2018.10.100

[16] M. Bal-Ghaoui, A. HEY, A. Jilbab, A. Bourouhou “Optimizing ultrasound image classification through transfer learning: fine-tuning strategies and classifier impact on pre-trained inner-layers”, *Informatyka Automatyka Pomiary Gospod Ochronie Srodowiska*, Vol. 13, No. 4, 2023. https://doi.org/10.35784/iapgos.4464

[17] L. Alkema, A. E. Raftery, P. Gerland, S. J. Clark F. Pelletier “Estimating trends in the total fertility rate with uncertainty using imperfect data: examples from West Africa”, *Demographic research*, Vol. 26, pp. 331-362, 2012. https://doi.org/10.4054/DemRes.2012.26.15

[18] M. Ala’raj, M. Majdalawieh, M. F. Abbod “Improving binary classification using filtering based on k-NN proximity graphs”, *Journal of Big Data*, Vol. 7, art. no. 15, 2020. https://doi.org/10.1186/s40537-020-00297-7

[19] D. A. Adeniyi, Z. Wei, Y. Yongquan “Automated web usage data mining and recommendation system using K-nearest neighbor (KNN) classification method”, *Applied Computing and Informatics*, Vol. 12, No. 1, pp. 90-108, 2016. https://doi.org/10.1016/j.aci.2014.10.001

[20] S. A. Abdullah “Hybrid model based on ReliefF algorithm and K-nearest neighbor for erythemato-squamous diseases forecasting”, *Arabian Journal for Science and Engineering*, Vol. 47, No. 2, pp. 1299-307, 2022. https://doi.org/10.1007/s13369-021-05921-z

[21] M. Wu, J. Zhou, Y. Peng, S. Wang, and Y. Zhang “Deep learning for image classification: a review”, *International Conference on Medical Imaging and Computer-Aided Diagnosis*, pp. 352-362, 2024. https://doi.org/10.1007/978-981-97-1335-6_31

[22] L. Wang, R. Tao, H. Hu, Y. R. Zeng “Effective wind power prediction using novel deep learning network: stacked independently recurrent autoencoder”, *Renewable Energy*, Vol. 164, pp. 642-655, 2021. https://doi.org/10.1016/j.renene.2020.09.108

[23] A. Sumathi, S. Meganathan, B. V. Ravisankar “An intelligent gestational diabetes diagnosis model using deep stacked autoencoder”, *Computers Materials and Continia*, Vol. 69, No. 3, pp. 3109-3126, 2021. https://doi.org/10.32604/cmc.2021.017612

[24] D. O'Neill, A. Lensen, B. Xue and M. Zhang, “Particle Swarm Optimisation for Feature Selection and Weighting in High-Dimensional Clustering”, *2018 IEEE Congress on Evolutionary Computation (CEC)*, Rio de Janeiro, Brazil, pp. 1-8, 2018. https://doi.org/10.1109/CEC.2018.8477974

[25] Y. V. S. Murthy, S. G. Koolagudi, T. K. J. Raja “Singer identification for Indian singers using convolutional neural networks”, *International Journal of Speech Technology*, Vol. 24, No. 3, pp. 781-796, 2021. https://doi.org/10.1007/s10772-021-09849-5

[26] C. Finn, P. Abbeel, S. Levine “Model-agnostic meta-learning for fast adaptation of deep networks”, *Proceedings of the 34th International Conference on Machine Learning*, Vol. 70, pp. 1126-1135, 2017. [CrossRef]

[27] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra and T. Lillicrap “Meta-learning with memory-augmented neural networks”, *Proceedings of the 33rd International Conference on Machine Learning*, Vol. 48, pp. 1842-1850, 2016. [CrossRef]

[28] H. J. Ye, L. Ming, D. C. Zhan and W. L. Chao, “Few-Shot Learning with a Strong Teacher”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 46, No. 3, pp. 1425-1440, 2024. https://doi.org/10.1109/TPAMI.2022.3160362

[29] X. Liu, C. Wang, J. Bai, G. Liao “Fine-tuning pre-trained convolutional neural networks for gastric precancerous disease classification on magnification narrow-band imaging images”, *Neurocomputing*, Vol. 392, pp. 253-267, 2020. https://doi.org/10.1016/j.neucom.2018.10.100

[30] B. Gunel, J. Du, A. Conneau, A. Stoyanov “Supervised contrastive learning for pre-trained language model fine-tuning”, *arXiv preprint arXiv*: 2011.01403, 2021. https://doi.org/10.48550/arXiv.2011.01403

[31] J. Pan “Feature-based transfer learning with real-world applications”, *Hong Kong University of Science and Technology*, 2010.

https://doi.org/10.14711/thesis-b1118218

[32] Y. Wang, S. Nazir, M. Shafiq "An overview on analyzing deep learning and transfer learning approaches for health monitoring", *Computational and Mathematical Methods in Medicine*, Vol. 2021, No. 1, art. no. 5552743, 2021. https://doi.org/10.1155/2021/5552743

[33] B. Tan, Y. Zhang, S. Pan, Q. Yang "Distant domain transfer learning", Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 31, No. 1, 2017. https://doi.org/10.1609/aaai.v31i1.10826

[34] M. H. Van den Berg, A. Overbeek, H. J. van der Pal, A. B. Versluys, D. Bresters, F. E. van Leeuwen, C. B. Lambalk, G. J. L Kaspers, and E V D D Broeder "Using web-based and paper-based questionnaires for collecting data on fertility issues among female childhood cancer survivors: differences in response characteristics", *Journal of medical Internet research*, Vol. 13, No. 3, art. no. e1707, 2011. https://doi.org/10.2196/jmir.1707

***