JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-20, No.-3, March (2025) pp 135 - 155 ISSN (Print) 0973-8975

MULTI-MODEL STACK ENSEMBLE DEEP LEARNING APPROACH FOR MULTI-DISEASE PREDICTION IN HEALTHCARE APPLICATION

Bhaskar Adepu[1], Dr. T. Archana[2]

[1]Department of CSE, Kakatiya University, Warangal, Telangana, India.

[2]Department of CSE, UCE, Kakatiya University, Bhadradri Kothagudem. Telangana, India.

E-mail: [1]ab.it@kitsw.ac.in, [2]archanapraneeth@gmail.com

Corresponding Email: **Bhaskar Adepu**

https://doi.org/10.26782/jmcms.2025.03.00009

(Received: November 07, 2024; Revised: February 24, 2025; Accepted: March 7, 2025)

Abstract

In the modern era of computers, numerous disciplines are witnessing the development of massive data volumes. Statistics are important in healthcare engineering because they provide insights into various diseases and match patient data. These datasets serve two functions: improving illness prediction and examining large data reservoirs to identify previously unknown disease patterns. Leveraging deep learning models, it becomes feasible to detect and forecast the early stages of numerous diseases based on individual health conditions. Nonetheless, the current landscape of illness prediction encounters several challenges such as inadequate large-scale datasets, logistical delays, the imperative for more precise and dependable predictions, and the intricacy of the models themselves. This paper introduces an innovative method for disease prediction utilizing deep learning, particularly focusing on an ensemble-based multi-disease prediction model. Datasets for lung cancer, cervical cancer, chronic renal disease, Parkinson's illness, and HCC survival are sourced from the trustworthy UCI repository for experimentation. A robust stacked deep ensemble model is proposed combining the InceptionResNetV2, EfficientNetV2L, and Xception architectures. This model integrates pre-processing techniques and employs the Tuna Swarm Optimization (TSO) Algorithm for feature selection in executing multi-label disease prediction. The suggested deep learning algorithms' performance is evaluated using criteria such as precision, specificity, sensitivity, accuracy, and error rate. This assessment demonstrates the potential of the recommended method to significantly contribute to the healthcare system by offering consistent and reliable predictions across various types of illnesses, as shown in a comparative analysis against existing models.

Keywords: Convolutional Deep Learning, Healthcare, Prediction, Stacking Ensemble Learning, Tuna Swarm Optimization.

I. Introduction

As society progresses, people's lifestyles and environments are constantly changing, leading to increased risks of various diseases. Major illnesses like brain health issues, cardiovascular diseases, vision problems, diabetes, and cancer have a significant impact worldwide, affecting human health and productivity. [V]. These diseases also add to societal pressures and raise medical costs. Disease prediction aims to foresee an individual's future disease risk. Doing this manually is not only difficult but also requires significant resources. [XXIII]. There is a strong desire for additional tools to improve disease forecasting and detection. Machine learning, which is one of several approaches to understanding complex data, has rapidly advanced, with deep learning emerging as the most prominent branch in the field of machine learning [XX]. Deep learning's ability to analyze complex data, extract key features from multidimensional data, efficiently handle unstructured data, and achieve higher-accuracy classification surpasses earlier technological methods. This also helps more people understand and engage with deep learning. [XXIV]. Deep learning has advanced significantly in recent years, leading to breakthroughs in areas like image and speech recognition. Its use in disease prediction has also become prominent, showing impressive results [XV]. Medical data works well with deep learning models, which can handle many tasks efficiently. In 2023, Khazaee Fadafen et al. [XVII], a hybrid architecture was presented that integrates dilated DTL processes and efficient learning to classify colorectal cancer (CRC) based on histopathology images. The proposed technique outperforms conventional deep learning models in both training and testing. Extensive trials using Kather_texture_2016 and NCT-CRC-HE-100K datasets indicate the proposed framework's generalizability. Furthermore, the hybrid deep learning method beats cutting-edge techniques in terms of computing efficiency and time. In 2023, Ismail et al. [XIV], introduced a novel approach for integrating MRI and PET images is proposed for diagnosing Alzheimer's disease (AD) across three patient groups: normal control, mild cognitive impairment, and AD. Additionally, the Multi-Objective Grasshopper Optimization Approach (MOGOA) is introduced as a multi-objective optimization method for optimizing the layers of the MultiAz-Net. The proposed deep ensemble model is evaluated on one multi-class categorization, three binary categorizations, and four AD classification tasks using a publicly available Alzheimer's neuroimaging dataset. The performance of the proposed approach is evaluated through multiple tests. In 2023, Anand, et al. [II], a weighted average ensemble deep learning model is proposed for brain tumour classification using grid search to combine optimal weight combinations from three feature spaces. The dataset comprises 3928 MRI images from 111 individuals with lower-grade gliomas. The ensemble model prevents overfitting by combining multiple models, outperforming individual models with accuracy, precision, and F1-scores of 97%, 97.25%, and 97.5%, respectively.

In 2022, Sagar Deep, et al. [III], the framework used an ensemble of four pre-trained DCNNs to extract features from Chest X-ray images, which are then classified using a fully connected layer. Results show that this multi-model ensemble outperforms a single classifier. In 2021, Ampavathi et al. [I] aimed to create a multi-disease prediction model using an improved deep learning method. The experiment employs datasets from

the UCI repository for diseases like lung cancer, Diabetes, Parkinson's disease, liver tumour, heart disease, Parkinson's disease, Alzheimer's disease and Hepatitis. Data is normalized and processed with a weight function using the JAMVO algorithm for feature extraction. Hybrid deep learning algorithms like RNN and DBN are used for prediction. In 2021, Hsu et al. [VIII], introduced a computer-assisted diagnostic system which is powered by intelligent learning models. The experiment employs datasets with missing data and erroneous information. The GA-CFS algorithms recommended analyzing all of the qualities as inputs to get the optimum solution. After that, the specific group is trained using various supervised classification approaches to determine how well the models perform. This method requires the availability of labelled data for training, which may not always be readily available or may be time-consuming and expensive to collect. In 2021, Dubey et al. [IV] proposed a Deep learning model used to forecast diseases. The sickness prediction was based on the datasets for AD, LC, PD, BC, TD, Dermatology, and heart disease. Medical datasets for various ailments were acquired from several sources. The L-BOA hybrid method was used to perform OFS over the full dataset's characteristic set. The original statistical features were computed and integrated with the most effectively extracted attributes. The L-BOA technique was utilized to maximize the hidden neuron count in both DBN and NN, using the identical feature set. In 2021, Essa et al. [VII], utilized deep learning to create a multi-model system for classifying ECG data. It includes two deep-learning bagging models and a refinement process to reduce false positives. This approach achieves a high overall accuracy of 95.81%, surpassing the best existing method by over 1%. It also outperforms other approaches in terms of average positive predictive value and F1 score. In 2020, Liu, et al. [XVIII], presented the MMEL-3DCNN, a multi-model ensemble learning architecture based on 3D convolutional neural networks. It reduces model training time and selects the best model for prediction based on nodule size. The approach was validated using the LIDC-IDRI dataset, and experiments confirmed its performance.

Khamparia, et al. [XVI], suggested a unique feature selection approach using deep learning to combine outputs from multiple classifiers. They use a genetic approach to improve prediction accuracy and reduce processing time. Experimental results show that their deep network ensemble technique outperforms existing classifiers and majority voting in accuracy and execution time. This approach also reduces generalizability error and helps classifiers learn nonlinear connections not possible with single-stage predictions. While there are various cancer prediction models, none are completely reliable, and each has its weaknesses. Combining multiple classification algorithms through stacking could improve performance over using just one model [XXV]. A multi-model ensemble is when several individual models' predictions are combined using another model. This second model learns how to best blend the initial predictions, resulting in a final set of predictions [XXII]. This study aims to combine three classification models—LeNet, AlexNet, and Xception—using deep neural networks to create a multi-model ensemble for predicting various illness states. A deep neural network will then merge the outputs of these three models to produce the final prediction [XIX]. The proposed technique was evaluated using datasets related to Parkinson's disease, lung cancer, chronic kidney disease, cervical cancer, and

hepatocellular carcinoma [XXI]. The findings demonstrate that the deep learning-based multi-model ensemble technique achieves notably high precision in its predictions with shorter training times, provides more accurate predictions compared to single classifiers, and utilizes limited clinical data more effectively [VI].

The paper includes the following sections. Section 2 reviews existing literature, setting the stage for the study. Section 3 introduces the recommended technique. Section 4 discusses the outcomes, comparing and analyzing the results. Finally, Section 5 concludes the paper, summarizing the findings and suggesting future research directions.

II. Literature review summary and Problem Statement

Predicting at an early stage is also a considerable expense. The multi-disease prediction model allows for the possibility of predicting multiple diseases. Therefore, the length of time is decreased and the possibility of death rate is also decreased. Hence, this study suggests employing a deep learning-based multimodal ensemble for multi-disease prediction. By leveraging limited clinical data, this approach yields more accurate predictions compared to single classifiers or majority voting methods. Additionally, it addresses overfitting concerns, streamlines processes, and reduces errors.

Configurations

- To address the problem of low accuracy and consistency in disease prediction, this work proposes creating a deep learning-based multimodal ensemble technique for predicting numerous diseases.
- To predict multiple diseases on various disease databases such as ‘‘LC, CC, PD, HCC, and CKD Survival datasets’’ are gathered from various sources.
- The Tuna Swarm Optimization (TSO) algorithm is employed in this study to achieve optimal feature selection, which is further enhanced through the use of a cross-validation approach.
- The individual’s classifiers DenseNet-121, AlexNet, and Xception are used in the classification process called the first stage.
- For multi-disease prediction proposed deep learning-based ensemble model using deep CNN is used to classify the first stage outcome which is called the second stage. The deep CNN network utilizes more accuracy to predict multi-disease prediction.

III. Methodology

The entire process of the proposed deep learning multi-model ensemble method is shown in Figures 1 and 2. In this article, five datasets of Lung cancer, Parkinson's disease, Hepatocellular Carcinoma dataset, chronic kidney disease, and Cervical cancer dataset are considered including gene samples and observations. The preprocessing technique uses normalization, Image scaling, Affine Distortions, and Elastic Distortion to improve the generalization and decrease the over-fitting. The project employs the Tuna Swarm optimization algorithm for feature selection and the cross-validation

technique to split the dataset into training and testing sets. Feature selection is crucial for overcoming challenges like overfitting caused by small sample sizes and high data dimensionality. Using ResNet50V2, Efficient Net B7, InceptionV3, VGG-16, and InceptionResNetV2 models as first-stage classifiers, the project aims to classify gene features. Subsequently, a convolutional deep neural network ensemble classifier combines the predictions from these models to enhance accuracy and reduce generalization error, enabling the application of deep neural networks and providing richer information for prediction.

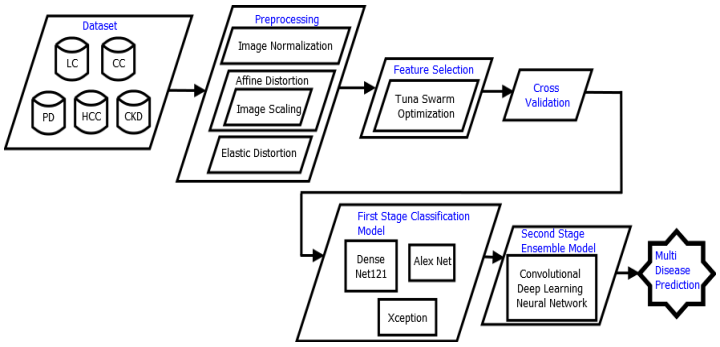

Fig. 1. Proposed Diagram

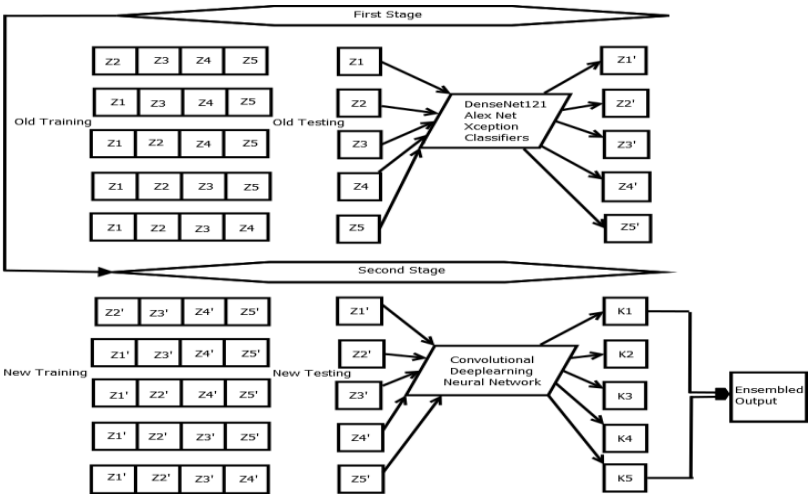

Fig. 2. Proposed deep learning-based ensemble method

Preprocessing

Image Normalization

Normalization adjusts pixel values in an image to a standard range, aiding in illumination adjustment.

Affine Distortions

Affine distortion applies displacement fields to pixels, improving generalization and reducing overfitting.

Image Scaling

Image scaling resamples and resizes images, using bicubic interpolation in this paper for better pixel value calculation.

Elastic Distortion

Elastic distortion expands datasets by generating random displacement fields and convolving them with a Gaussian function, improving generalization in deep neural networks.

Feature Selection and Cross-validation

Tuna swarm optimization (TSO) Algorithm

Inspiration.

Tuna, known scientifically as Thunnini, are carnivorous marine fish that exhibit continuous swimming behaviour. They are top marine predators, using their intelligence to efficiently forage in groups. This study introduces Tuna Swarm Optimization (TSO), a novel optimization technique inspired by tuna's natural foraging behaviours. The mathematical Model is provided in this section, we present a detailed description of the mathematical model underlying the proposed algorithm. We outline the key equations and principles that govern the algorithm's behaviour and explain how these components interact to achieve the desired optimization outcomes.

Initialization Process

In the Tuna Swarm Optimization (TSO) algorithm, which is similar to other swarm-based meta-heuristics, the optimization process begins by randomly generating initial populations distributed evenly across the search space. This approach ensures that the algorithm explores a wide range of potential solutions from the start, facilitating a more comprehensive search for optimal solutions.

$$Z_x^{int} = rand.(vb - hb);\ x = 1,2,\dots\dots,nb \tag{1}$$

Z_x^{int} represents the x^{th} starting individual, vb and hb are the upper and lower limits. The variable nb indicates the number of tuna populations, and $rand$ is a random value uniformly distributed between 0 and 1. These parameters play a crucial role in the Tuna Swarm Optimization (TSO) algorithm, influencing the initial distribution of tuna populations within the search space.

Spiral Foraging

Small schooling fish-like herring and sardines use a dense, constantly changing formation to evade predators, while tuna employ a tightly spiralling group strategy to chase prey. Tuna share knowledge by following the fish in front of them, facilitating

coordination within the school. This behaviour inspires the mathematical formula for the spiral foraging strategy.

$$Z_x^{d+1} = \begin{cases} \alpha_1 \cdot \left(Z_{best}^d + \beta \cdot \left|Z_{best}^d - Z_x^d\right|\right) + \alpha_2 \cdot Z_x^d; & x = 1 \\ \alpha_1 \cdot \left(Z_{best}^d + \beta \cdot \left|Z_{best}^d - Z_x^d\right|\right) + \alpha_2 \cdot Z_{x-1}^d; & x = 2,3,\ldots.., np, \end{cases} \quad (2)$$

$$\alpha_1 = k + (1-k) \cdot \frac{d}{d_{max}} \quad (3)$$

$$\alpha_2 = (1-k) - (1-k) \cdot \frac{d}{d_{max}} \quad (4)$$

$$\beta = e^{bh} \cdot \cos(2\pi b) \quad (5)$$

$$l = e^{3cos}\left(\left(\left(d_{max} + {}^1/_d\right) - 1\right)\pi\right) \quad (6)$$

$Z_x^{d+1}$is d+1 iteration of x^{th}individual. The equation incorporates several variables: Z_{best}^d stands for the current optimal individual (food), while $\alpha_1$and α_2 are weights that control how much individuals move towards the best option and the previous choice. The constant "k" decides how closely tuna follow these options at the start. "d" is the current iteration number, and d is the maximum number of iterations, ranging from 0 to 1.

When all tuna spiral around food, they efficiently search the immediate area. But if the best tuna can't find food, blindly following it doesn't help. To solve this, we suggest adding a random search location. This lets each tuna explore more, improving the Tuna Swarm Optimization's ability to explore widely.

$$Z_x^{d+1} = \begin{cases} \alpha_1 \cdot \left(Z_{rand}^d + \beta \cdot \left|Z_{rand}^d - Z_x^d\right|\right) + \alpha_2 \cdot Z_x^d; & x = 1 \\ \alpha_1 \cdot \left(Z_{rand}^d + \beta \cdot \left|Z_{rand}^d - Z_x^d\right|\right) + \alpha_2 \cdot Z_{x-1}^d; & x = 2,3,\ldots.., np, \end{cases} \quad (7)$$

A random location Z_{rand}^d is selected within the search space to serve as a reference point for exploration. For example, Meta-heuristic algorithms typically prioritize extensive global exploration initially, gradually shifting towards more targeted local exploitation. In TSO, as the number of iterations increases, the reference points for spiral foraging transition from random to optimal positions.

$$Z_x^{d+1} = \begin{cases} \alpha_1 \cdot \left(Z_{rand}^d + \beta \cdot \left|Z_{rand}^d - Z_x^d\right|\right) + \alpha_2 \cdot Z_x^d; & x = 1,\ if\ rand < \frac{d}{d_{max}} \\ \alpha_1 \cdot \left(Z_{rand}^d + \beta \cdot \left|Z_{rand}^d - Z_x^d\right|\right) + \alpha_2 \cdot Z_{x-1}^d; & x = 2,3,..,np \\ \alpha_1 \cdot \left(Z_{best}^d + \beta \cdot \left|Z_{best}^d - Z_{best}^d\right|\right) + \alpha_2 \cdot Z_x^d; & x = 1,\ if\ rand \geq \frac{d}{d_{max}} \\ \alpha_1 \cdot \left(Z_{bset}^d + \beta \cdot \left|Z_{best}^d - Z_{best}^d\right|\right) + \alpha_2 \cdot Z_{x-1}^d; & x = 2,3,\ldots.,np, \end{cases} \quad (8)$$

Parabolic Foraging.

Tunas feed in both spiral and parabolic patterns. Tuna exhibit two concurrent behaviours in their search for food: they form a parabolic shape centered around the food as a reference point, and they scan their surroundings. These two procedures are carried out simultaneously, with each procedure having a selection probability of 50%. The mathematical model describing this behaviour is as follows,

$$Z_x^{d+1} = \begin{cases} Z_{best}^d + rand \cdot \left(Z_{best}^d - Z_x^d\right) + TF' \cdot p^2 \cdot \left(Z_{best}^d - Z_x^d\right), & if\ rand\ < 0.5, \\ TF' \cdot p^2 \cdot Z_x^d, & if\ rand\ \geq 0.5, \end{cases} \quad (9)$$

$$p' = \left(1 - \frac{d}{d_{max}}\right)^{\left(d/d_{max}\right)} \quad (10)$$

Where TF' is a random variable with a value of either 1 or -1.

In the Tuna Search Optimization (TSO) process, individuals use two foraging tactics to locate prey. They start randomly in the search space and, during each cycle, randomly choose a tactic or regenerate their location based on a probability distribution. The parameter "s" is adjusted dynamically. The process continues until a termination condition is met, returning the optimal individual and its fitness value.

Cross-validation method

Cross-validation is utilized in this study to combat overfitting and enhance generalization. A fivefold cross-validation technique is adopted, where the dataset is divided into five partitions. Four partitions are used for training, while one is reserved for testing. This process is iterated over all possible combinations, and the model's performance is averaged across different partitions.

Classification methods

DenseNet-121

The denseNet-121 model of densely linked convolutional networks, published in late 2017, is made up of networks with blocks, each with direct access to the gradient of the error function, cost function, and the block's starting input. Of course, there are also convolutional and pooling layers.

AlexNet Architecture

AlexNet uses five convolutional layers (CONV1-5) and two fully connected (FC) layers (FC6 and FC7). We convert the two-dimensional MIAS dataset to three dimensions for AlexNet. We extract 4096 features from each image using the FC6 and FC7 layers for high accuracy.

Xception

Xception is a successful pre-trained algorithm based on the ImageNet dataset. It uses a modified convolution technique and has 36 layers for feature learning, achieving better results and faster training times.

Multi-model ensemble-driven Deep Neural Network

Multi-model ensemble techniques aim to enhance prediction accuracy by combining forecasts from various models. In this particular study, a convolutional deep neural network (CNN) is employed as the ensemble model, which integrates different classifiers to predict sickness more accurately. This approach leverages the strengths of each individual classifier, resulting in improved overall performance in sickness prediction. A neural network's purpose is to use nonlinearity to generate output from a

large number of input signals. A neural network must be a nonlinear approximation with input, output, and multiple hidden layers.

Convolutional layer

Convolution is used to extract features by sliding filters over the input in separate convolution layers to build a corrected activation map based on feature identification via edges, corners, and so on. The resulting output feature map includes many convolutions, signified by the stated formula provided in (Eq. 11) as follows:

$$s_x^f = h\left(\sum_{y \in K_x} s_y^{f-1} * z_{xy}^f + l_x^f\right) \tag{11}$$

wherez_{xy} represents the kernel, f represents the layer, l_x represents the bias, and K_x represents the collection of input maps that will be utilized to generate the output s_x.

Max pooling layers

The subsampling layers, also known as pooling layers, play a crucial role in reducing variance by computing feature values over different arrays of gene samples. These layers accept an input volume of size $E_1 * Q_1 * T_1$ and utilize two key parameters, spatial extent (H) and stride (P), to generate an output volume of size $E_2 * Q_2 * T_2$.

$$E_2 = (E_1 - H)/P + 1, R_2 = (R_1 - H)/P + 1 \; and \; T_2 = T_1 \tag{12}$$

The utilization of max and average pooling in this distribution is advantageous as it speeds up the convergence of convolutional neural networks (CNNs) and improves the network's ability to generalize.

SoftMax layer

SoftMax regression is a form of logistic regression that generalizes the binary outcome to multiple classes. It calculates the probabilities of each class being the correct one using a SoftMax function and then predicts the class with the highest probability. This makes it suitable for multi-class classification problems.

Backpropagation gradient descent training algorithm

Backpropagation is crucial for training CNNs. It involves both forward and backward passes. In multi-class classification, the goal is to minimize the mean squared error over training samples by comparing the network's output to the actual class labels and adjusting the errors through backpropagation.

$$W = \frac{1}{s}\sum_{s=1}^{s}\sum_{r=1}^{u}(d_r^s - j_r^s)^2 \tag{13}$$

Where d_r^sc represents the response of r^{th}output in the current layer to the s^{th}input pattern.

The output vector of the current layer y with input layer q is given by,

$$i^y = h(E^y i^{y-q} + l^y) \tag{14}$$

Here l^y is the generated bias and $E^y i^{y-q}$is the summing junction output.

The backpropagation algorithm is used to update the weights of a neural network by propagating the error from higher layers to lower layers. This adjustment minimizes the loss function and enhances the model's performance.

$$\delta^{y} = (E^{y+1})^{D}\delta^{y+1} \odot h'(E^{y}i^{y-q} + l^{y}) \quad (15)$$

The weight update, denoted by ΔE^{y}, was accomplished using the formula shown below:

$$\frac{\delta W}{\delta E^{y}} = i^{y-q}(\delta^{y})^{D} \quad (16)$$

$$\Delta E^{y} = -\mu \frac{\delta W}{\delta E^{y}} \quad (17)$$

In this work, we combine multiple models using convolutional deep learning to enhance the classification performance. In the first stage, the supplied dataset Z was partitioned into eight different subsets: $Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7, Z_8$, where $Z_x \in \{a_x, b_y\}$ contains various labelled points. In the first round, $Z_2 V Z_3 V Z_4, V Z_5 V Z_6 V Z_7, V Z_8$ are utilized as the training set, whereas $Z_1 \in a_x, b_y$ is the test set. The dataset is divided into eight subsets to perform fivefold cross-validation, ensuring a robust evaluation of the model. , i.e., $Z_1', Z_2', Z_3', Z_4', Z_5', Z_6', Z_7', Z_8'$ where $Z_x' = [N_x, j_y]$, where x=1, 2, 3,….., 8. This approach allows for a comprehensive analysis of the dataset, enhancing the model's ability to generalize to unseen data. Our goal is to categorize different neuromuscular diseases using deep learning. We pick input neurons reflecting sample attributes and use hidden layers to improve classification. Instead of focusing solely on the output layer, we prioritize the average value for better results. In the second stage, we use deep learning to establish complex relationships and make accurate predictions.

IV. Result and Discussion

Dataset

The detailed description of the dataset used for the evaluation of the performance based on the evaluation metrics considered is given in detail in this section. The dataset link is provided in the reference section [IX – XII].

Lung cancer

The information utilized in this study is a set of data related to lung cancer, which was initially published and subsequently provided in the UCI machine learning repository with a name such as "Lung Cancer Data set". The potential of the optimal discriminant plane in challenging scenarios was demonstrated using this dataset. The information in this dataset pertains to the various pathological types of lung cancer. There are 56 elements used to observe three types of lung cancer, with a total of 32 observations. The dataset provides information on three types of lung cancers that are commonly found.

Parkinson's disease

The PD dataset includes audio samples previously used in other studies to identify Parkinson's disease, obtained from UCI's ML repositories. The study, conducted at

Istanbul University's Cerrahpasa Faculty of Medicine, involved 186 individuals with PD (106 males, 80 females, aged 32 to 86) and 62 healthy participants (22 males, 40 females, aged 40 to 81). Recordings were made at a frequency of 44.1 kHz, with each person recording three sets of vowel sounds after medical exams.

Hepatocellular Carcinoma dataset

The dataset about hepatocellular carcinoma (HCC) comprises data from 165 individuals diagnosed with this condition, sourced from a University Hospital in Portugal. The dataset contains a variety of demographic, laboratory, overall survival features, and risk factors. There are 49 features in total, selected based on the EASL-EORTC Clinical Practice Guidelines, which are considered the most advanced guidelines for managing HCC. The dataset is composed of both quantitative and qualitative variables, totalling 23 and 26 respectively. Most of the dataset contains incomplete information, with a missing data rate of 10.22%. Furthermore, only a tiny percentage of patients (4.85%) had comprehensive information across all areas. The one-year survival rate is the objective variable, and it is represented by a binary variable: 0 (death) or 1 (survival).

Chronic Kidney Disease dataset

This research work utilizes the CKD dataset. This dataset has also been utilized by numerous researchers. The UC Irvine Machine Learning Repository is furnishing this dataset, which can be accessed on the UCI website. The dataset comprises 400 examples and 24 features, including 1 target attribute. The two-class labelling of the target attribute represents either CKD or non-CKD. In 2015, data was gathered from different hospitals.

Cervical cancer

The database "Cervical Cancer Risk Factors for Biopsy" was provided by the UCI repository. There is data on the actions, characteristics, and health records of 858 individuals in the compilation. The dataset for hospital patients has some missing values because some patients elected not to answer specific questions for privacy reasons. There are 858 items in the collection, and each one has 32 characteristics. There are 32 variables and medical records of 858 female patients included in the dataset. The data comprises 32 variables and the medical records of 858 women patients, which consist of details like STDs, smoking habits, IUD, age, and other similar factors.

Experimental Setup

The full set of programming instructions was written in Python within a Jupyter Notebook.3.2 Dataset Collection. The performance of the suggested technique is assessed in trials with five different datasets. The datasets are briefly described.

Parkinson's disease

The dataset contains 31 individuals, 23 of whom have been diagnosed with Parkinson's disease, and the measures acquired are relevant to biomedical voice analysis. The data

differentiates between healthy individuals (0) and those with Parkinson's disease (1) based on the "status" column.

Lung cancer

Hong and Young utilized the optimal discriminant plane to exhibit its efficacy in situations with ambiguous conditions. The data pertaining to three distinct types of pathological lung cancers is presented. The data set comprises class and nominal predictive attributes with integer values ranging from 0 to 3.

Chronic Kidney Disease

CKD is a condition where kidney function declines gradually, leading to waste buildup in the body, which can result in a wide range of issues. The CKD dataset comprises various features such as red blood cells, albumin, age, blood pressure, sugar, and other related attributes. The purpose of the dataset is to use the provided attributes to predict whether a patient has CKD or not.

Cervical cancer

CC is cancer that develops in the cervix, the lower part of the uterus that connects to the vagina. The cervical cancer dataset comprises variables such as age, count of sexual partners, count of pregnancies, and additional attributes. The objective of the dataset is to predict whether an individual has cervical cancer or not, utilizing the given attributes. The dataset contains details concerning the survival rates of individuals diagnosed with cervical cancer.

HCC Survival Data Set

HCC refers to a form of liver cancer. The HCC Survival Data Set comprises variables such as age, gender, tumour size, and additional factors. The objective of the database is to forecast the likelihood of survival among individuals diagnosed with HCC, utilizing the given attributes. The dataset contains details regarding the survival rates of individuals diagnosed with HCC.

Evaluation and Analysis

This section discusses research performed on various datasets to categorize multiple diseases. A description of the evaluation metrics and the obtained results is provided. The results are compared with previous studies in this final sub-section.

Performance Metrics

When assessing the effectiveness of a deep learning model, various performance metrics are commonly employed. These metrics provide insights into how well the model fits the data and can help in fine-tuning its parameters to improve efficiency. Some of the key performance metrics used in this experiment include:

Accuracy tells us how often the model is correct. Precision focuses on the proportion of correctly predicted positive cases. Recall looks at the proportion of actual positive cases that were predicted correctly. Specificity considers the proportion of actual negative cases that were predicted correctly. The F1 Score balances precision and recall. AUC measures the overall performance of the model. Error Rate tells us how often the model is wrong.

Evaluating these metrics helps researchers and practitioners understand how well a deep learning model performs and allows them to improve its effectiveness.

$$A_{ccuracy} = \frac{(T_rp+T_rn)}{(T_rp+T_rn+F_ap+F_an)} \quad (18)$$

$$P_{recision} = \frac{T_rp}{T_rp+T_rp} \quad (19)$$

$$F1_{Score} = \frac{S_e \cdot p_r}{p_r+S_e} \quad (20)$$

$$S_{pecificity} = \frac{T_rn}{F_an} \quad (21)$$

$$R_{ecall} = \frac{Tp'}{Tp'+Fn'} * 100 \quad (22)$$

$$E_{rror} = \frac{Fp'+Tn'}{Tp'+Tn'+Fp'+Fn'} * 100 \quad (23)$$

Analysis of Classification Performance

Compare the predictive accuracy, precision, recall, specificity, F1 score, AUC, and Error of each method such us ResNet50V2, Efficient Net B7, Inception V3, VGG-16, Inception ResNet V2, and Ensemble comparing two methods on five different datasets: Cervical cancer, Lung cancer, Hepatocellular Carcinoma, Chronic Kidney Disease, and Parkinson's disease. We're using graphs to show how these methods perform differently as we change the discrimination threshold.

Table 1: Result of Predictive Accuracy

Methods	ResNet50V2	Efficient Net B7	Inception V3	VGG - 16	Inception ResNet V2	Ensemble Method
LC	97.3	95.8	96.2	97.0	95.8	99.5
CC	95.7	96.4	95.1	97.5	96.4	99.6
PD	95.4	94.9	96.8	96.2	97.4	99.7
HCC	96.5	97.5	98.5	98.3	97.2	99.6
CKD	96.8	95.9	94.8	97.5	96.8	99.4

The results in Table 1 show the prediction accuracy of different methods for each cancer dataset. Our ensemble method outperforms single classifiers and the majority voting algorithm, demonstrating more stable performance across the datasets.

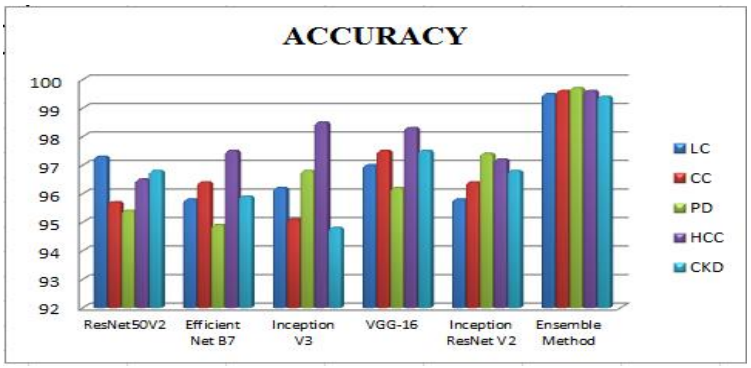

Fig. 3. Comparison Analysis of Predictive Accuracy

Figure 3 presents a comparison analysis of Accuracy. Table 1 shows that our method is better than majority voting for predicting cancer types, with higher accuracy across all datasets 99.5%, 99.6%, 99.7%, 99.6%, and 99.4% for the LC, CC, PD, HCC, and CKD datasets, respectively. In another hand the individual classifier of ResNet50V2, Efficient Net B7, Inception V3, VGG-16 and Inception ResNet V2 have the accuracy rate for LC dataset is 97.3%, 95.8%, 96.2%, 97.0% and 95.8% respectively, then accuracy rate for CC dataset is 95.7%, 96.4%, 95.1%, 97.5% and 96.4% respectively, then PD dataset is 95.4, 94.9, 96.8, 96.2 and 97.4 respectively, accuracy for HCC is 96.5%, 97.5%, 98.5%, 98.3% and 97.2% respectively and predictive accuracy of CKD is 96.8%, 95.9%, 94.8%, 97.5% and 96.8% respectively.

Table 2: Results of Predictive Precision

Methods	ResNet50V2	Efficient Net B7	Inception V3	VGG-16	Inception ResNet V2	Ensemble Method
LC	98.7	94.7	97.9	96.4	93.9	99.2
CC	96.8	93.7	94.4	95.2	97.3	99.0
PD	96.4	95.7	94.8	98.6	97.3	99.6
HCC	95.9	96.4	98.5	97.6	97.4	99.2
CKD	97.8	96.3	97.2	96.9	95.1	98.9

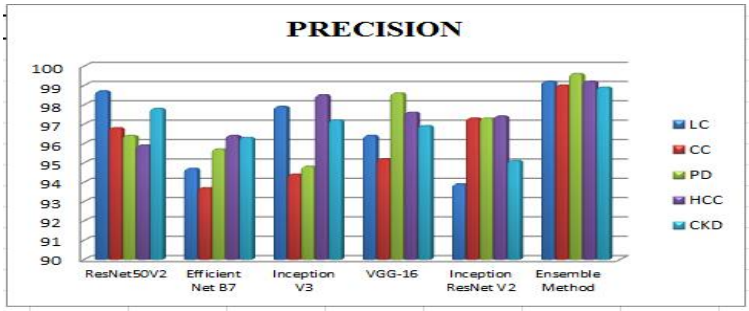

Fig. 4. Comparison analysis of Precision

Figure 4 depicts a precision evaluation for five cancer data sets: LC, CC, PD, HCC, and CKD. Table 2 shows the results of predictive Precision and the data suggests that combining classifiers improves precision, leading to better classification performance compared to using each classifier alone. Our deep learning-based method outperforms majority voting in all five datasets, achieving precision rates of 99.2%, 99%, 99.6%, 99.2%, and 98.9%, respectively.

Table 3: Results of Predictive Recall

Methods	ResNet50V2	Efficient Net B7	Inception V3	VGG-16	Inception ResNet V2	Ensemble Method
LC	97.6	96.2	95.9	96.1	98.6	99.0
CC	94.9	97.6	93.9	95.7	97.4	99.1
PD	94.6	93.8	95.8	97.4	96.2	98.4
HCC	97.7	96.4	98.3	97.2	95.7	99.8
CKD	95.8	94.2	96.4	93.9	96.5	98.6

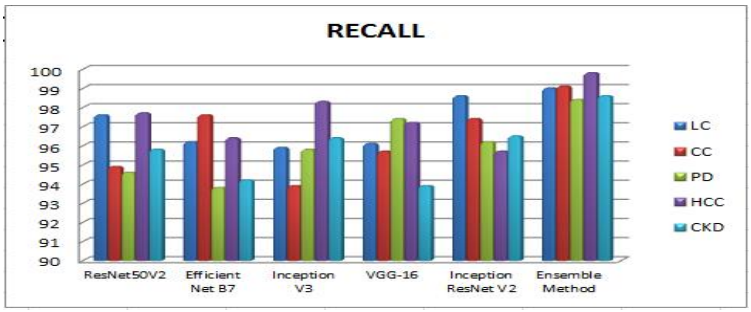

Fig. 5. Comparison Analysis of Recall

A comparison of recall for other individual classifiers including ResNet50V2, Efficient Net B7, Inception V3, VGG-16, and Inception ResNet V2 reveals that the recall values for these approaches for LC, CC, PD, HCC, and CKD 99.0%, 99.1%, 98.4%, 99.8% and 98.6% respectively. This graphical model is presented in Figure 5. The proposed ensemble models achieve high recall values compared to the other single classifiers. The proposed strategy results in other current methodologies concerning performance as displayed in Table 3.

Table 4: Results of Specificity

Methods	ResNet50V2	Efficient Net B7	Inception V3	VGG-16	Inception ResNet V2	Ensemble Method
LC	98.4	95.7	97.4	96.7	95.9	98.9
CC	97.6	95.4	94.9	91.2	96.5	98.5
PD	97.4	98.2	96.9	97.5	96.8	98.1
HCC	94.8	97.8	95.5	93.9	97.5	98.2
CKD	89.7	95.6	94.8	97.2	96.9	99.3

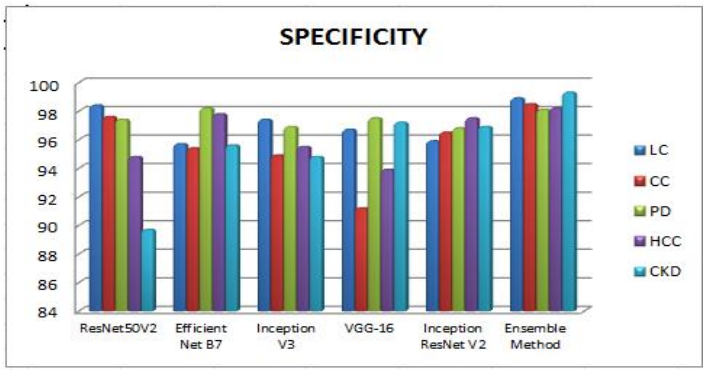

Fig. 6. Comparison analysis of Specificity

Figure 6 shows comparisons between the proposed model and five individual classifiers and the proposed multi-model ensemble approach has a high specificity value for various datasets of LC, CC, PD, HCC, and CKD of 98.9%, 98.5%, 98.1%, 98.2%, and 99.3% respectively, while other recommended classifiers have minimum specificity values for proposed datasets. The ensemble method achieved the maximum

value of specificity for five given datasets. The result of specificity is denoted in Table 4.

Table 5: Results of F1 Score

Methods	ResNet50V2	Efficient Net B7	Inception V3	VGG-16	Inception ResNet V2	Ensemble Method
LC	97.8	97.6	96.8	97.5	98.4	99.3
CC	94.3	95.9	93.8	96.7	97.4	98.9
PD	96.4	95.3	94.9	97.2	96.9	98.9
HCC	97.5	94.6	95.8	97.5	96.9	98.6
CKD	93.6	95.7	97.4	94.9	95.9	98.3

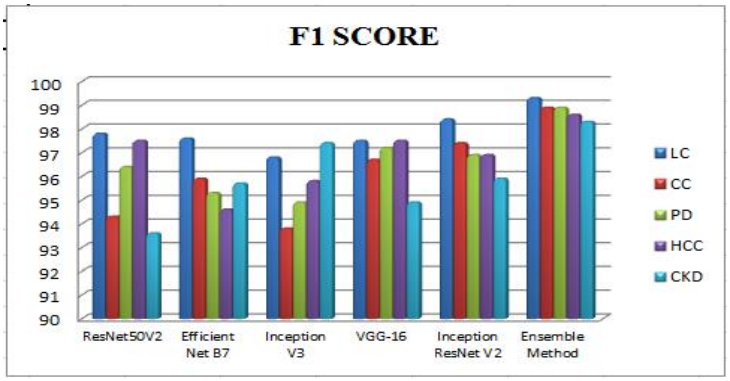

Fig. 7. Comparison analysis of F1 Score

Figure 7 shows that comparison analysis of F1Score and the numerical results of F1 score is presented in Table 5. The proposed ensemble model achieved more F1score value compared to the other five individual classifiers for LC (99.3%), CC (98.9%) PD (98.9%) HCC (98.6%) and CKD (98.6%), The proposed datasets LC (97.8%), CC (94.3%), PD (96.4%), HCC (97.5%) and CKD (93.6%) achieve F1score for ResNet50V2 classifier, Datasets LC (97.6%), CC (95.9%), PD (95.3%), HCC (94.6%) and CKD (95.7%) for Efficient Net B7, datasets LC (96.8%), CC (93.8%), PD (94.9%), HCC (95.8%) and CKD (97.4%) for Inception V3, datasets LC (97.5%) ,CC (96.7%) ,PD (97.2%) ,HCC (97.5%) and CKD (94.9%) for VGG-16 and datasets LC (98.4%), CC (97.4%), PD (96.9%), HCC (96.9%) and CKD (95.9%) for Inception ResNet V2. The result of classification analysis ensemble method performance is better than other individual classifier methods. Table 6 shows the outcomes of F1score values.

Table 6: Results of AUC

Methods	ResNet50V2	Efficient Net B7	Inception V3	VGG-16	Inception ResNet V2	Ensemble Method
LC	95.6	94.9	96.2	98.6	97.8	98.2
CC	95.8	97.5	96.2	97.7	96.8	98.2
PD	96.9	94.9	95.6	96.2	96.8	99.6
HCC	95.2	97.4	98.1	95.2	97.9	98.7
CKD	95.8	97.8	94.7	97.4	98.2	99.2

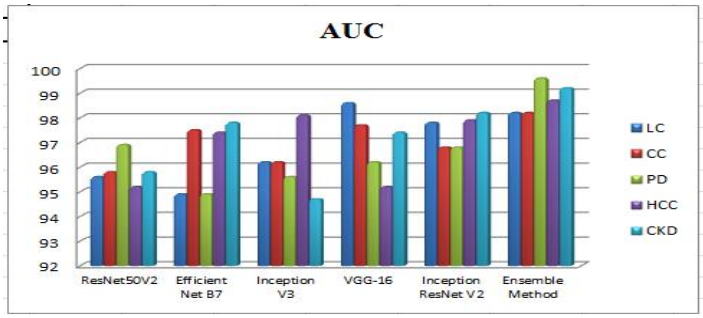

Fig. 8. Comparison analysis of AUC

Table 6 displays the sensitivity rates, which are viewed in Figure 8 for evaluating the outcome was performed suggested method compared to previous techniques. The suggested Model demonstrates better AUC compared to other classifier techniques (ResNet50V2, Efficient Net B7, Inception V3, VGG-16, and Inception ResNet V2) with five datasets such as LC, CC, PD, HCC, and CKD and achieving 98.2%, 98.2%,99.6%, 98.7% and 99.2% respectively.

Table 7: Result of Error

Methods	ResNet50V2	Efficient Net B7	Inception V3	VGG-16	Inception ResNet V2	Ensemble Method
LC	0.015	0.022	0.033	0.023	0.020	0.001
CC	0.031	0.026	0.011	0.032	0.017	0.001
PD	0.017	0.004	0.013	0.024	0.008	0.000
HCC	0.006	0.0017	0.008	0.023	0.010	0.000
CKD	0.027	0.034	0.024	0.009	0.016	0.006

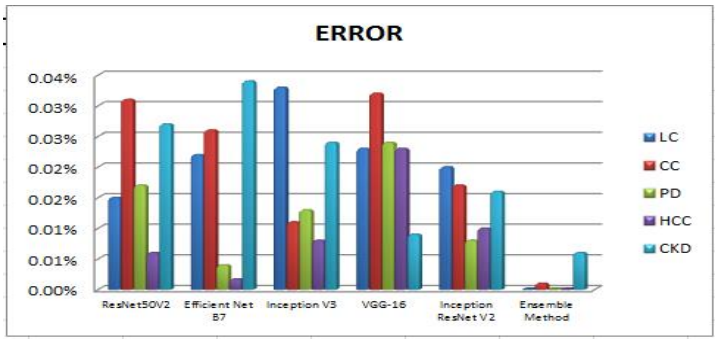

Fig. 9. Comparison Analysis of Error

Figure 9 illustrates the Error analysis comparing the proposed and other various classifier approaches. The suggested strategy has an error (0.00%) of LC, CC, PD, HCC, and CKD for ResNet50V2, Efficient Net B7, Inception V3, VGG-16, and Inception ResNet V2 classifiers. The error rate is better than other classifiers. Table 7 denotes the result of the error rate. The suggested model outperformed other techniques in terms of results.

Analysis of Convergence computational time

The time a technology takes to complete a task is called computational time. Table 8 shows the computational time of the suggested and current models. Analyzing the data shows that the proposed classification ensemble approach takes LC (1.02%), CC (1.55%), PD (2.88%) HCC (1.32%), and CKD (1.59%) sec and outperforms other classifiers have high computational time that value is shown in Figure 10. The advised strategy worked because running time was lowered. The proposed study is suitable for multi-disease classification. The figure depicts computational time with Ensemble methods and individual classifier methods.

Table 8: Evaluation of Computational Time

Methods	ResNet50V2	Efficient Net B7	Inception V3	VGG-16	Inception ResNet V2	Ensemble Method
LC	3.48	4.89	2.78	6.13	8.67	1.02
CC	2.76	3.64	4.69	4.86	4.65	1.55
PD	5.96	5.58	7.97	3.76	3.44	2.88
HCC	1.87	4.63	3.11	5.98	2.61	1.32
CKD	4.76	2.16	4.65	7.02	5.24	1.59

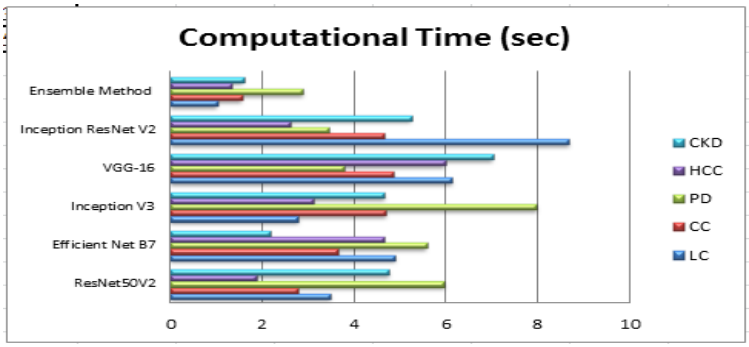

Fig. 10. Graphical design of computational time (sec)

V. Conclusion

This study undertakes a thorough comparative analysis of various deep-learning models applied to disease prediction, focusing on five distinct datasets: Lung cancer, Cervical cancer, Parkinson's disease, Hepatocellular Carcinoma, and chronic kidney disease. Our evaluation criteria encompass a comprehensive set of metrics, including predictive accuracy, precision, recall, specificity, area under the curve (AUC), F1 score and error rates. The deep learning models under investigation include ResNet50V2, Efficient Net B7, Inception V3, VGG-16, and Inception ResNet V2. Additionally, ensemble methods such as majority voting and our proposed ensemble method are explored. Through meticulous scrutiny of each method's performance across these datasets, our objective is to identify the most effective approach for disease prediction. Utilizing graphical representations, we present a comparative analysis that elucidates the strengths and weaknesses of each model, thereby offering valuable insights for the advancement of predictive modelling in healthcare engineering. By focusing on Techniques such as attention visualization, feature importance analysis, and model-

agnostic explanations that can help elucidate the model's decision-making process areas of enhancement, researchers can drive forward the development of more accurate, reliable, and interpretable disease prediction models, ultimately leading to improved healthcare outcomes.

Conflict of interest

The authors declare no conflict of interest available for this manuscript.

References

I. Ampavathi, Anusha and VijayaSaradhi, T., : 'Multi disease-prediction framework using hybrid deep learning: an optimal prediction model.' *Computer Methods in Biomechanics and Biomedical Engineering*. Vol. 24(10), pp. 1146-1168, 2021. 10.1080/10255842.2020.1869726

II. Anand, Vatsala, et al., : 'Weighted average ensemble deep learning model for stratification of brain tumor in MRI images.' *Diagnostics*. Vol. 13(7), pp. 1320, 2023. 10.3390/diagnostics13071320

III. Deb, Sagar Deep, et al., : 'A multi model ensemble based deep convolution neural network structure for detection of COVID19.' *Biomedical signal processing and control*. Vol. 71, pp. 103126, 2022. 10.1016/j.bspc.2021.103126

IV. Dubey, A.K., : 'Optimized hybrid learning for multi disease prediction enabled by lion with butterfly optimization algorithm.' *Sādhanā*. Vol. 46(2), pp. 63, 2021. 10.1007/s12046-021-01574-8

V. Dwight L Evans, Dennis S Charney, : 'A journal of psychiatric neuroscienceand therapeutics "Mood disorders and medical illness: a major public health problem.' *A journal of psychiatric neuroscience therapeutics*. Vol. 54(13), pp. 177-180, 2003. 10.1016/S0006-3223(03)00639-5.

VI. Ehab, E., Xianghua, X., : 'An Ensemble of Deep Learning-Based Multi-Model for ECG Heartbeat Arrhythmia classification.' *IEEE*. Vol. 9, 2021. 10.1109/ACCESS.2021.3098986

VII. Essa, Ehab, and XianghuaXie, : 'An ensemble of deep learning-based multi-model for ECG heartbeats arrhythmia classification.' *IEEE access*, Vol. 9, pp. 103452-103464, 2021. 10.1109/ACCESS.2021.3098986

VIII. Hsu, C.H., Chen, X., Lin, W., Jiang, C., Zhang, Y., Hao, Z & Chung, Y.C.,: 'Effective multiple cancer disease diagnosis frameworks for improved healthcare using machine learning.' *Measurement*. Vol. 175, pp. 109145, 2021. 10.1016/j.measurement.2021.109145

IX. https://archive.ics.uci.edu/dataset/174/parkinsons

X. https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease

XI. https://archive.ics.uci.edu/dataset/383/cervical+cancer+risk+factors

XII. https://archive.ics.uci.edu/dataset/423/hcc+survival

XIII. https://archive.ics.uci.edu/dataset/62/lung+cancer

XIV. Ismail, Walaa, N., FathimathulRajeena, P.P and Mona, Ali, A.S., : 'A meta-heuristic multi-objective optimization method for alzheimer's disease detection based on multi-modal data.' *Mathematics*. Vol. 11(4), pp. 957, 2023. 10.3390/math11040957

XV. Kevin Zhou, S., Hayit Greenspan, Christos Davatzikos, James Duncan, S., Bram Van Ginnek, : 'A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends.' *Case Studies With Progress Highlights, and Future Promises*. Vol. 109, pp. 5, 2021. 10.1109/JPROC.2021.3054390

XVI. Khamparia, Aditya, et al., : 'A novel deep learning-based multi-model ensemble method for the prediction of neuromuscular disorders.' *Neural computing and applications.* Vol. 32, pp. 11083-11095, 2020. 10.1007/s00521-018-3896-0

XVII. KhazaeeFadafen, Masoud, and KhosroRezaee, : 'Ensemble-based multi-tissue classification approach of colorectal cancer histology images using a novel hybrid deep learning framework.' *Scientific Reports*. Vol. 13(1), pp. 8823, 2023. 10.1038/s41598-023-35431-x

XVIII. Liu, Hong, et al., : 'Multi-model ensemble learning architecture based on 3D CNN for lung nodule malignancy suspiciousness classification.' *Journal of Digital Imaging*. Vol. 33, pp. 1242-1256, 2020. 10.1109/ACCESS.2021.3098986

XIX. Saurabh, A., Arya, K.V., Meena, Y.K., : 'MultiFusionNet: Multilayer Multimodal Fusion of Deep Neural Networks for Chest X-Ray Image Classification.' 10.48550/arXiv.2401.00728

XX. Shaveta, D., Kumar, M., Ayyagari, M.R., Kumar, G., : 'A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning.' *Archives of Computational Methods in Engineering*. Vol. 27, pp. 1071-1092, 2022. 10.1007/s11831-019-09344-w

XXI. Wenjie Kang, Lan Lin, Baiwen Zhang, XiaoqiShen, ShuicaiWu, : 'Multi-model and multi-slice ensemble learning architecture based on 2D convolutional neural networks for Alzheimer's disease diagnosis.' *Computers in Biology and Medicine*. Vol. 136, 2021. 104678. 10.1016/j.compbiomed.2021.104678

XXII. Yawen, X., Jun, W., Zongil, L., Xiaodong, Z., : 'A deep learning-based multi-model ensemble method for cancer prediction.' *Computer methods and programin Biomedicine*. Vol. 153, pp. 1-9, 2018. 10.1016/j.cmpb.2017.09.005

XXIII. Yogesh, K.D., ElviraI Smagilova, Gert Aarts, et al., : 'Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy.' *International Journal of Information Management*. Vol. 57, 2021. 10.1016/j.ijinfomgt.2019.08.002

XXIV. Yuyang He, You Zhou, Tao Wen, Shuang Zhang, Fang Huang, XinyuZou, Xiaogang Ma, YueqinZhu, : 'A review of machine learning in geochemistry and cosmochemistry: Method improvements and applications.' *Applied Geochemistry*. Vol. 140, pp. 105273, 2022. 10.1016/j.apgeochem.2022.105273

XXV. YuyanWang, Dujuan Wang, Na Geng, YanzhangWang, Y. Yunqiang, J. Yaochu, : 'Stacking-based ensemble learning of decision trees for interpretable prostate cancer.' *Applied Soft Computing*. Vol. 77, pp. 188-204, 2019. 10.1016/j.asoc.2019.01.015