International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

An Elevated Cancer Detection Methodology Using a Hybrid Transfer Learning Approach for Classifying Microarray Cancer Data

Swati Sucharita[1*], Barnali Sahu[2], Tripti Swarnkar[3]

Submitted: 25/05/2023 **Revised**: 17/07/2023 **Accepted**: 28/07/2023

Abstract: Cancer has emerged as a prominent issue over the last decade, necessitating timely identification for effective treatment. Consequently, the development of an automated diagnostic system for precise cancer detection has considerable significance. The analysis of microarray data is a significant challenge due to the complexity of the data, limited sample size, uneven distribution of classes, the presence of noisy data structure, and more variability in feature values. The existing Machine Learning (ML) led to lesser classification accuracy and its training is proven to be quite expensive with over-fitting problem. In contrast to the drawbacks of conventional machine learning (ML) methods like support vector machines, decision trees, logistic regression, etc., deep learning (DL) methods provide a variety of benefits including large data compatibility, automated feature engineering, and ease of use. The objective of this study is to propose the development of Mote Carlo Relief-F feature filtering (MCRelief-F) as a feature estimator that can effectively provide quality assessments of features while dealing with intricate situations characterized by significant interdependencies among features. When it comes to feature selection, SOT (scyphozoan optimization technique) is more successful than traditional approaches because it has better convergence ability, search stability, and optimum-seeking ability. The Hybrid Extensive Kernel Convolutional Neural Network (HEKCNN-LSTM-TL) uses an extensive convolution kernel for local convolution and long-term memory (LSTM) with transfer learning to improve classification accuracy while shortening training times. The suggested technique is tested on three common microarray cancer datasets, including brain, breast, and leukemia, and the feature values are scaled using artificial SOT. Classification accuracy, precision, f-measure, specificity, sensitivity, and MCC are used to assess how well the given strategy performs. A comparison of the suggested approach's performance with state-of-the-art approaches is done, and the results show that it performs better than many of the current methods, notably on the leukemia dataset.

Keywords: *Cancer, microarray data, Machine Learning, Deep Learning, feature filtering, feature selection convolutional neural network, long-term memory, Transfer Learning and classification.*

1. Introduction

Cancer is one of the deadliest illnesses. Early detection and treatment of cancer is essential [1]. One of the most often discovered cancers is lung cancer, which has been exceeded in prevalence by occurrences of female breast cancer. Cancer cases and fatalities in 2020 are shown in Fig. 1. In the first stages of development, instances are found in two-thirds of cases [2]. DNA microarray data may classify and identify gene expression for cancer subtype diagnosis and prognosis. AI-based learning algorithms, which are essential tools, are the most efficient way to gather the most important components of gene expression data and a fundamental element of gene classification. This article aims to provide a comprehensive analysis of many tactics discussed in existing literature, along with an overview of the datasets used for implementing these approaches. Additionally, the article will explore the advantages and disadvantages connected with the application of these strategies [2]. Deep learning has emerged as a fundamental technique in clinical cancer research, including well-established variants such as Convolution Neural Networks,

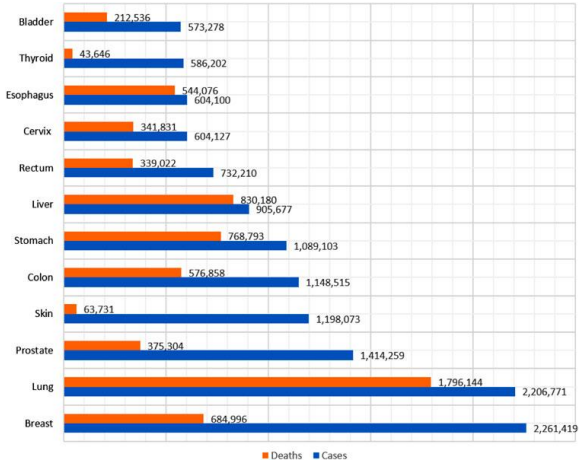

Fig.1. 2020 cancer death and cases

Artificial Neural Networks, and Autoencoders. These variations play a crucial role in informing decision-making processes related to illness diagnosis and therapeutic interventions. Over the course of time, the complexity and challenges associated with the identification, analysis, and

[1,2] *Department of Computer Science and Engineering,*
Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, India
[3]*Department of Computer Application*
Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, India
[1*]*swatisucharita08@gmail.com*

treatment of illnesses, including cancer, have progressively intensified. The investigation of cancer is a significant subject of inquiry within the field of medicine. Furthermore, it is worth noting that medical datasets often exhibit noise, changes in feature values, and uneven class distributions. These characteristics might lead to overfitting issues and thus degrade the accuracy of classification models [3]. Research in the area of microarray data analysis is required, particularly in the context of cancer classification, serves to ascertain and comprehend the characteristics that contribute to the progression of cancer. Finding genes that contribute to certain biological outcomes and using those genes to predict additional observations are important functions of the microarray data categorization technique. Early identification of cancer has a crucial role in facilitating the formulation of treatment strategies by subject specialists, hence improving the overall survival rate of those affected by cancer [4]. Therefore, the issue necessitates the careful development of a model that can effectively analyze input patterns representing objects and accurately forecast the category of the item in question. Consequently, it is essential to create a prediction model that can provide precise results when applied to the provided test data [5].

The classification of microarray cancer data encompasses many key processes, including data collection from the original source, pre-processing of the data, feature selection, classification of the data, and subsequent analysis after the classification process. Feature selection is a crucial procedure that involves the identification and selection of significant genes from a vast pool of highly correlated and informative genes. This decision refines data components for a classifier, improving classification accuracy [6]. The process of feature selection is of utmost importance in the classification of cancer data as it enables the identification of an optimum and relevant subset of characteristics. This, in turn, leads to an improvement in classification accuracy and computational stability [7]. The taxonomy of dimensionality reduction methods is shown diagrammatically. Five methods for feature selection, the employment of ensemble, filter, wrapper, embedding, hybrid, and newly developed methods have been fully explored, with datasets utilized and accuracy results reported. Each technique's benefits and drawbacks have also been discussed [8]. In order to get more knowledge about the condition, preparing decisive actions and improving the cancer diagnostic process eventually depend on the analysis of microarray cancer data [9]. Recent efforts proposing neural network models for issues relating to cancer prediction are discussed in [10]. To test the suggested models, all the publications taken into consideration employed gene expression datasets.

To categorize microarray cancer data in this study, the authors suggest using a deep learning-based classifier [11]. To understand how features, behave during training and predict the class of unobserved data, deep learning uses a vast amount of data. Three common microarray cancer datasets, including brain, breast, and leukaemia diseases, were taken into consideration in order to verify the suggested methodology. This paper addresses the problem of medical diagnosis by presenting an intelligent approach for cancer detection based on microarray datasets. A microarray dataset often contains a relatively high number of characteristics compared to a small number of samples. The following is the work's primary contribution:

- This research suggested using gene expression patterns to categorize cancers.
- The MCRelief-F and SOT algorithms are used in the proposed system to filter significant genes (features) from the input matrix and select features, respectively.
- The categorization of microarray data for cancer detection is then done using HEKCNN-LSTM-TL.

Following is an organization of the remaining portion of the paper. Please see Section 2 for a discussion of relevant studies. In Section 3, the approach is given. Section 4 discusses the experiment and outcomes. Section 5 discusses and compares, and Section 6 concludes and suggests more studies.

2. Related Work

Gene expression profiles were used in all of the earlier research described below to categorize cancer using a variety of techniques. Kar et al., [12] used an adaptive K-nearest neighbour (KNN)-based gene selection technique and the particle swarm optimization (PSO) to locate a few acceptable genes adequate for categorization. The small round blue cell tumor (SRBCT), acute lymphoblastic and acute myeloid leukemia (ALL), and mixed-lineage leukemia (MLL) datasets are three benchmark microarray datasets on which the suggested method of identifying the smallest relevant collection of genes is tested. In terms of classification accuracy on blind test samples, number of informative genes, and computation time, the suggested approach is effective. However, various computer approaches have failed to identify a few key genes due to the limited number of relevant samples and the large number of genes in microarray data.

Mishra & Bhoi [13] analyzed the use of the improved ANFIS (EANFIS) technique to categorize cancer genes. Since the convergence time of ANFIS rises throughout the course of learning, the Manta ray foraging optimization (MaFO) approach is a hybrid to improve classification performance overall. The first stage of the classification process begins with a pre-filtering step using the Ensemble

Kalman Filter (EnKF) technique. Adaptive density-based spatial clustering with noise (ADBSCAN) clusters pre-processed genes with comparable characteristics. However, these methods continue to fall short of expectations in terms of performance. Halder & Kumar [14] built a novel active learning method, ALRFC (active learning with rough-fuzzy classifier) to classify cancer samples. The suggested method may manage the ambiguity, overlapping, and indistinct gene expression data subtypes. However, this approach is often restricted since it tends to rely on professional evaluation to detect tumors and because it may be difficult to distinguish between various cancer subtypes.

To choose characteristics and categorize cancer, relaxed Lasso and support vector machines (rL-SVM) were suggested by Mazlan et al., [15]. In order to compete with other approaches already in use, the 10-fold cross-validation increase classification accuracy for all datasets. The experimental results in this work show that the suggested strategy produces superior classification accuracy while using fewer chosen feature genes. High-dimension and small-sample cancer data may be classified using rL-SVM in big datasets. The high dimensionality, noise, and irrelevant genes of this microarray, however, are only a few of the issues it has.

Shafi et al., [16] integrated "Mean Decrease Accuracy" and "Mean Decrease Gini" as feature selection methodologies inside the well-recognized Random Forest classifier might alleviate the challenges posed by high-dimensional data and expedite computational processes. Moreover, this amalgamation has the potential to enhance the accuracy of the prediction model. Furthermore, we've shown a comparison between a model with feature selection and one without. The comprehensive experimental findings have shown that the suggested feature-selection model is advantageous and successful and that it outperforms other models in terms of accuracy.

Hambali et al., [17] employed Information Gain - Modified Bat Algorithm (InfoGain-MBA) features selection technique has been presented for identifying relevant and educational characteristics from highly dimensional Microarray cancer datasets. To assess the technique, four classifiers were employed: Classification and regression tree (CART), Random Forest, and C4.5. The findings obtained demonstrate the potential of the suggested technique for the categorization of microarray cancer data. In each of the seven employed datasets, the random forest has a 100% accuracy rate with few genes. The most effective threshold for each of the datasets was also determined via further research. Finding meaningful genes from the hundreds of accessible genes in the microarray data is one of the study's main limitations.

Alrefai et al., [18] developed a comprehensive system for classifying cancer. Three tasks make up this framework. Firstly, the high dimensionality of the microarray dataset is reduced through particle swarm optimization with ensemble learning (PSO-ensemble). Additionally, the Adaptive Self-Training Method (ASTM) is used to resolve little difficulties. The categorization process was then completed using a convolutional neural network (CNN). CNN can identify the intricate non-linear correlations between characteristics and choose the most educational ones. Because transfer learning may shorten training periods and simplify computing tasks, it was used after CNN to incorporate categorization. Six microarray datasets are employed for the liver, breast, colon, prostate, central nervous system, and lung.

Alrefai & Ibrahim [19] adopted the ensemble learning approach and the suggested particle swarm optimization For feature selection and cancer classification. The proposed microarray-based cancer classification method outperforms most state-of-the-art methods and dominates the baseline ensemble method with 12% enhancement, with accuracy results of 100%, 92.86%, 86.36%, 100%, and 85.71% for leukemia, colon, breast, ovarian, and central nervous system, respectively. Nevertheless, these methods have not been effective in identifying the most relevant and informatic genes because of the low sample numbers in microarray data relative to the high dimensionality.

Khaire & Dhanalakshmi [20] created Improved Adam (iAdam) technique combines the Look-Ahead Mechanism with Adaptive Learning Rate for Each Parameter. When the gradient is calculated after the current speed, the traditional Adam's momentum increases. Additionally, Adam serves as a correctional component for Adam's forward velocity. Additionally, it works well for the high-dimensional dataset and converges to minimal error within the given epochs even at higher learning rates. When employed in the categorization of data sets independent of the research they were created from, signatures from several studies often exhibit poor resilience.

Zhang et al., [21] combined the principal component analysis method with an autoencoder neural network to give a framework for learning unsupervised features from gene expression patterns. An ensemble classifier using the AdaBoost algorithm (PCA-AE-Ada) predicted breast cancer clinical outcomes using the characteristics collected. To serve as a benchmark for the suggested approach throughout the trials, we created a second classifier using the identical classifier learning strategy (PCA-Ada), with the only variation being the training inputs.

Inference: Summarizing the preceding comments on gene expression data-based cancer prediction methodologies Achieving high prediction accuracy with few training samples, dealing with the dimensionality curse, and dealing with overlapping, indistinguishable, and complicated cancer subtype groups are all difficulties. This paper proposes a unique deep-learning cancer classification approach to overcome the aforementioned difficulties and the disadvantages of active learning algorithms used to assess gene expression data.

3. Proposed Methodology

Experimental datasets will be used in this part to benchmark the performance of the suggested work. In addition to HEKCNN-LSTM-TL, the section describes a data input representation system. Microarray data X are structured as matrix vectors and utilized for classification before HEKCNN-LSTM-TL is used. Researchers have concentrated on hybrid methods that combine CNN, LSTM, and TL since single Machine Learning (ML) techniques do not seem to be able to provide the best outcomes in terms of both prediction performance and stability when applied to cancer classification. The feature selection process involves using the feature subset generated by the MCRelief-F and SOT algorithms. The use of heatmaps facilitates the visualization of microarray data or data obtained from future sequencing research, such as bacterial analysis. Using a system of color-coding to represent the various values found in a matrix, a heatmap is a graphical depiction of data. The input of the picture into the HEKCNN-LSTM-TL program produced heatmaps showing feature selection power at each percentile of a certain feature set. Following that, classification is carried out using a subset of the SOT algorithm's characteristics. Based on gene expression data, the classification is carried out using the HEKCNN-LSTM-TL classifier. In Fig. 2, the suggested HEKCNN-LSTM-TL cancer detection system's general structure is shown.

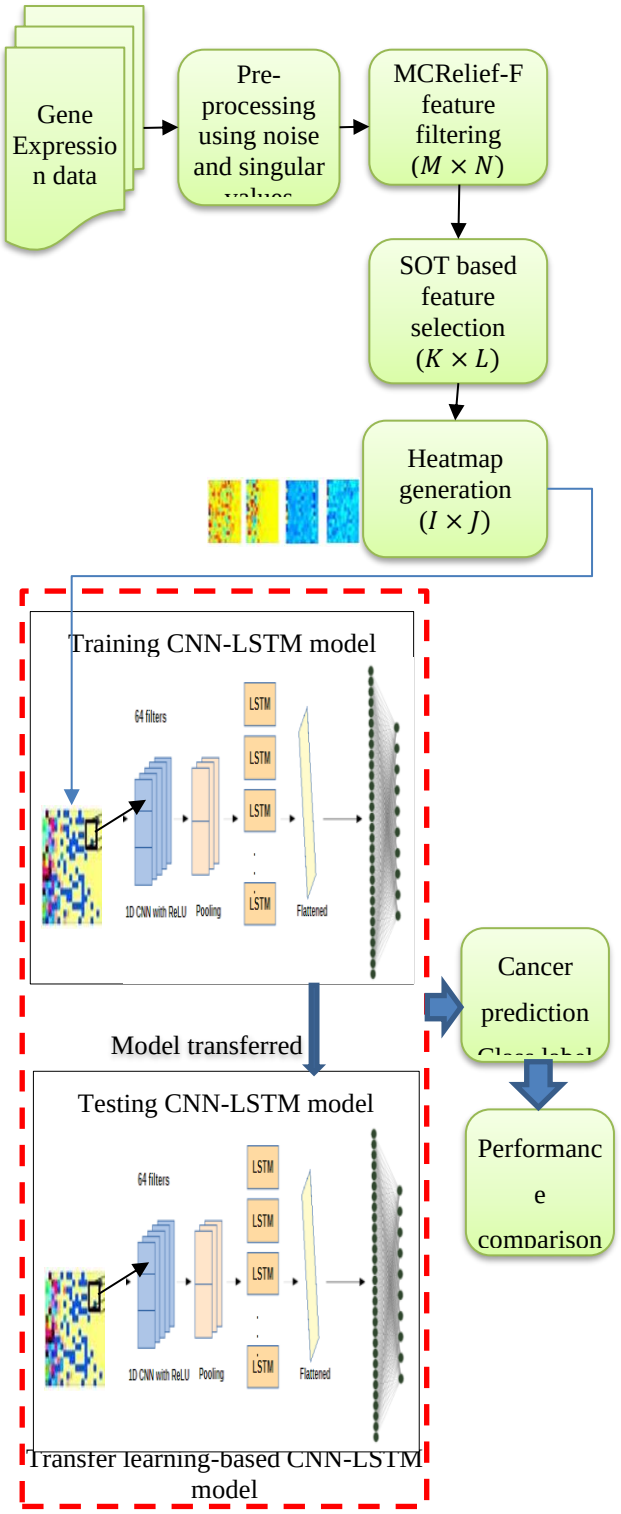

Fig.2. The proposed classification scheme for microarray data

Micro array dataset description

Classification accuracy has been used a lot to evaluate the success of the presented approach, which is the benchmark in disease diagnosis, on three datasets (brain, breast, and leukemia cancer).

Brain cancer gene expression dataset [22]: CSV file with 130 samples (rows) and 54676 genes' levels of gene expression in each sample. There are 4 different types of

brain cancer (plus healthy tissue) represented in this dataset (column "type").

Breast cancer gene expression dataset [23]**:** CSV file containing the gene expression levels of 54676 genes (columns) from 151 samples (rows). There are 5 different types of breast cancer (plus healthy tissue) represented in this dataset (column "type").

Leukemia gene expression – CuMiDa [24]**:** CSV file with 64 samples (rows) with 22284 genes' levels of gene expression in each sample. There are 5 different types of leukemia represented in this dataset (column "type").

MCRelief-F algorithm for Filtering Process

The Relief-F algorithm was combined with the MC approach in an effort to enhance its performance. This study was motivated by the previous work in [25]. Show how the feature selection issue for high dimensional data may be resolved using the MC approach in this section. Given that the challenge of feature selection is manifestly distinct from sequential decision-making problems, for which the MC algorithm was first developed, in this methodology, In the beginning, the feature selection issue is rephrased as a sequential choice problem. Each node of the search tree corresponds to a different characteristic when it is constructed. The action set is the set of features from the root node to the current node. A sequential decision-making issue is thus created out of the feature selection problem.

A search must be guided toward a better route via an evaluation function, much as in the MC technique. Relief-F's fundamental method involves selecting instances at random, calculating their closest neighbors, and adjusting a feature weighting vector to provide greater weight to qualities that distinguish the instance from its neighbors of other classes. It specifically looks for a reliable estimate of the chance that Pr to determine what each feature's weight should be $\mathfrak{F}$.

$$\mathfrak{w}\mathfrak{F} = Pr\,(DV\ of\ \mathfrak{F}|DC) - Pr(DV\ of\ \mathfrak{F}|SC) \qquad (1)$$

Where DV is the difference value, DC is different class and SC is same class. This strategy has performed well in a number of fields. One may get a feature weight vector similar to that in Eq. (2) by using the Relief-F technique to determine the feature weights of each feature in the feature set;

$$\mathfrak{w}\mathfrak{F} = [\mathfrak{w}\mathfrak{F}(1), \mathfrak{w}\mathfrak{F}(2), \dots, \mathfrak{w}\mathfrak{F}(m), \dots, \mathfrak{w}\mathfrak{F}(n)] \qquad (2)$$

where n the number of features, and m is how many characteristics the MC algorithm has eliminated before the simulation phase begins. Eq. (1) may be used to determine the weight of each feature in the vector. The finest stratagem fs^* the MC method generates a feature set that maximizes the sum of feature weights:

$$fs^* = \arg\max \sum_{i=1}^{m} \mathfrak{w}\mathfrak{F}(fs) \qquad (3)$$

In contrast to the actual decision-making issue, the same activities are not permitted in this feature reduction problem. Identical actions (features) are not allowed in either of the action (feature) sets, which means that the two action (feature) sets cannot be the same. With the help of the Upper Confidence Bounds (UCB) procedure, the actions (features) are sequentially selected, starting with the root node, which is initially an empty set.

$$R^* = \arg\max\left(\mathfrak{C}(s,R) + \mathrm{tc}\sqrt{\log\frac{N(s)}{N(s,R)}}\right) \qquad (4)$$

The reduction (feature) denoted as R* represents the child node located below the current node that possesses the highest UCB value. In this context, N(s) refers to the frequency of selection of the current node, N(s, R) represents the frequency of selection of the corresponding child action of the current node, and tc denotes a constant used for tuning. There must be a required equilibrium between the first (extraction) and second (exploration) sides of the UCB equation. The exploration term's contribution reduces when each node is visited as a result of the increasing denominator. On the other side, if the parent node visits a different child, the numerator rises, increasing the exploration values of siblings who haven't been visited yet. According to Browne et al. (2012), the exploration term's tc constant may be changed to either increase or decrease the quantity of exploration done. As more iterations are added, the search tree grows asymmetrically, and features with small classification weights (often referred to as noise features) are filtered out. This is how the UCB algorithm described above handles feature reduction.

Feature selection using SOT

Attribute elimination in the feature selection (FS) procedure may assist in determining the amount of the data, limiting dimensionality, reducing computing time and requirements, and enhancing performance prediction. Additionally, the choice of characteristics aids the prediction models in finding subtle nuances that might enhance the performance of the particular domain under consideration. Due to their effectiveness in reducing computational demands, increasing classification accuracy, reducing storage requirements, and solving complex optimization problems quickly, emerging technologies like meta- and hyper-heuristics optimization methods offer a new paradigm for FS. To discover the best feature set that maximizes classification performance, SOT is used in this study for this reason, as shown in Fig. 3. The SOT algorithm replicates the natural method used by jellyfish to find food in water currents [26].

Initialization of the scyphozoans: The initialization process is described in this section for the position of the scyphozoans, which is constructed based on the logistics map, where SX_i is the location of the i-th scyphozoan and is the chaotic logistic value. Using SX_0, which contains parameters (0, 1) and ϑ a parameter value of 4, the first population of Scyphozoa was produced.

$$X_{i+1} = \vartheta * SX_i(1 - X_i), 0 \leq SX_0 \leq 1 \quad (5)$$

In (16), where r is a random number, δ is a distribution coefficient larger than 0, and SX_{cur} is the best scyphozoan location right now, sea currents are formed. The symbol avg represents the average position of all scyphozoan.

$$SX_i(t+1) = (SX_i(t) + \mathbb{r}(0,1) \times SX_{cur}) - (\delta \times \mathbb{r}(0,1) \times \text{avg}) + SX_i(t) \quad (6)$$

Position updating based on scyphozoan movement: With the VM's movement as a guide, this section explains position updates. Both passive and aggressive motions regulate the scyphozoan's mobility within the swarm. The scyphozoan, for instance, just moves in situ when it exhibits passive motion. (6) shows the newly discovered scyphozoan location based on passive motion. When a search space has upper and lower bounds, u and l, respectively The coefficient of motion, m, which is expressed by a number larger than zero, measures the length of motion around the scyphozoan location. The successful utilization of local search space, however, is thought to occur when there is active mobility. The formula presented in (7) is used to define it.

$$SX_i(t+1) = SX_i(t) + m \times \mathbb{r}(0,1) \times (\mathbb{u} - \mathbb{l}) \quad (7)$$

$$SX_i(t+1) = \mathbb{u} * SX_i(t) + \mathbb{r}(0,1) \times \mathbb{l} * \overrightarrow{swim(SX)} \quad (8)$$

$$\overrightarrow{swim(SX)} = \begin{cases} X_j(t) - X_i(t) & if\ OF(SX_i) \geq OF(SX_j) \\ X_i(t) - X_j(t) & if\ OF(SX_i) \geq OF(SX_j) \end{cases} \quad (9)$$

Where, $\overrightarrow{swim}$ Depending on density value, each scyphozoan swims in the direction that will help it locate the most food. This behavior is represented in equation (5), where OF is the objective function depending on the location′s Griewank SX [30].

Process of time control mechanism: Scyphozoan movement caused by variations in sea current temperature, passive motions, and active movements are modeled using the time control mechanism. The movement in (10 is simulated). Scyphozoan swarms grow throughout time (via iteration). By making both aggressive and passive motions, each scyphozoan moves through the swarm to find a better location. Exploitation, where $exploit(t)$ changes with each repetition, is the term used to describe this behavior. The two required parameters in the proposed method are the number of swarms and $\max it$. where $\max iter$ is the number of iterations that are permitted in total.

$$exploit(t) = \left(1 - \frac{1}{\max it}\right) \times \delta(2\mathbb{L}(0,1) - 1) \quad (10)$$

Apply large rank value: This paper suggests a large rank rate ($\mathbb{L}$) strategy that is simple to use for combinatorial issues of illness detection and classification to move the scyphozoan vector in the journey sequence. The entire distribution cost is calculated using it and is shown in (11). On the basis of Objective Function (OF), which is primarily dependent on accuracy value, the mathematical formula for the feature selection issue is presented:

$$\min \sum_{i=1}^{n} OF(i) \quad (11)$$

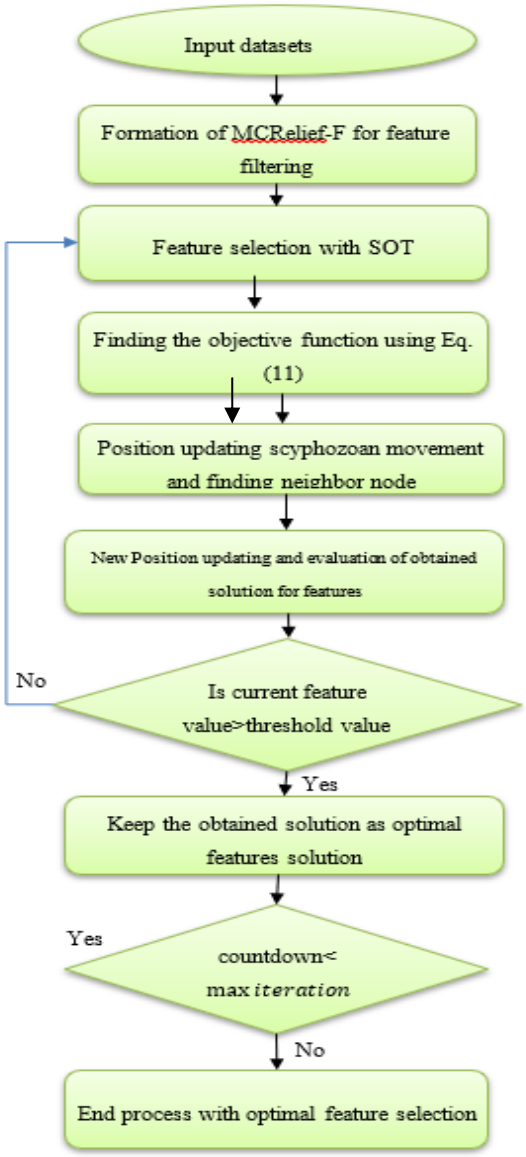

Fig.3. Flowchart for the SOT-based feature selection that has been proposed

Apply neighborhood exchange: Neighborhood discussion is used to enhance the scyphozoan algorithm's performance with each iteration. In this research, switch, and neighborhood jesting are two exchange rules that are suggested. Two randomly selected scyphozoan position vectors are subject to the swap rule, which reverses their locations. To illustrate the swap rule, two randomly chosen scyphozoan position vectors are switched. Scyphozoan and iterations have been set at 30 and 50, respectively. A collection of scyphozoan positions is transformed into a journey sequence using the L technique in each iteration. To find the optimal allocation solution in the current iteration, the previous solution is contrasted with the new Neighborhood exchange result. As a consequence, classification is accomplished with reduced energy consumption depending on the population and iteration parameters.

Only when the data center's capacity exceeds the existing accuracy value of a certain feature will a choice be made in order to ensure system stability by reducing the number of migrations. Based on the accuracy, a threshold value (a value ranging from 0 to 1) is chosen to help discover the best feature. The fundamental concept is to keep sending characteristics to the scyphozoan up until it becomes overwhelming, that is until the load on that scyphozoan surpasses the threshold value. Here, transmit the remaining characteristics to underloaded Scyphozoans that can be located using the random search approach rather than sending them to the overloaded machine.

The proposed HEKCNN-LSTM-TL model for cancer detection

In this part, a hybrid neural network incorporating Extensive kernel CNN and LSTM is proposed as a diagnostic model to increase the accuracy of cancer detection with limited effective data. Three components make up the model: a one-dimensional CNN component, an LSTM component, and a complete connection layer component. In order to lower the input data's dimension and extract its local spatial properties, the one-dimensional CNN portion will first be used. To overcome the delayed LSTM training issue brought on by data dimension explosion, after dimensionality reduction, the LSTM component will further extract the sequential features of the encoding characteristics. The fully linked layer component is then assigned to the encoding characteristics produced by the LSTM section. The diagnostic model may extract low- and medium-frequency characteristics from input data using fewer parameters due to the CNN component's convolutional layer's large convolution kernel for local convolution. The suggested hybrid neural network model's transfer learning technique improves the diagnostic model's domain flexibility. In this context, the transfer of valuable information pertaining to the intended work, acquired via extensive data training, enables the achievement of accurate diagnostic performance even when confronted with limited data availability. Consequently, there is no need to reconstruct the diagnostic model in response to changes in operational circumstances. The transfer learning technique used in this study offers greater benefits in addressing the variable working circumstances in engineering practice across domains than the previous data-driven deep learning methods.

Heatmap: For the two-dimensional representation of data matrix values, a heatmap is a visual depiction of each matrix member using colors. Little rectangles or pixels with a dark hue are used to represent high numbers in this example, whereas bright colors are used to represent little values. Heatmaps have a wide range of uses since they are good for visualization and may be colored in a number of ways depending on the application. Examples include studies of web page visits, the display of data, and the analysis of gene expression data. The feature vector acquired by SOT in this work was transformed into the image data required by HEKCNN-LSTM-TL using a heatmap.

CNN: The input layer, the hidden layer, and the latent layers are the three main components. Three distinct latent (hidden) layer types may be found: a fully connected layer, a pooling layer, or a convolutional layer [27].

EKConvolution layer: Using filters and the input data as a source, the convolution layer of an EKCNN extracts the features according to the specified dimension. Weights are chosen at random in this case. These weights are used to create a new feature map from the data when the convolutional kernel procedure with a m×n filter is applied. For neural networks that use two-dimensional images as input, this technique expands the network depth and enhances network characterization. It also reduces the network's weight parameters to prevent overfitting until the dataset is complete. The acquired dataset is subjected to the application of the Rectified Linear Unit (ReLU) activation function, which serves as a threshold function. Subsequently, the normalizing procedure is used to equilibrate the distribution of data, which has the potential to undergo alterations subsequent to the convolution process.

Max pooling layer: The pooling procedures are carried out in the output layer after the normalizing phase to provide improved feature maps. Pooling procedures are often carried out to reduce the amount of input in the subsequent convolution phase. The pooling procedure is carried out using the pool size and a stride value of two in the majority of pooling algorithms. The neuron dropout strategy is used during the training phase to prevent the overfitting issue during training. Neuron dropout is

calculated as 0.2 in value. For establishing connections with the neurons of the preceding layer and carrying out classification with the completely connected layer, the density of the neuron is set to 1024.

Softmax layer: To increase classification accuracy, the probability-based softmax function is utilized last. The output values are normalized using the softmax algorithm before being converted to probability values. Finally, based on these probabilities, the test data is categorized.

Long short-term memory network: As illustrated in Fig. 4, LSTM may solve classic RNNs' gradient disappearance and gradient explosion problems when dealing with long-term dependant information. These two issues arise in conventional RNNs as a consequence of the gradient being continuously attenuated or boosted throughout the error back propagation (BP) process, which causes gradient expansion or gradient disappearance. In the LSTM architecture, a single LSTM memory unit is designed to address the aforementioned issues. This unit has three essential components, namely a forgetting gate, an input gate, and an output gate. These gates serve the purpose of filtering, updating, and reporting information, respectively. This configuration is seen in Figure 4. The gate structure has several concealed neurons, and the faults may be disseminated uniformly over these concealed neurons. As a result, the network's gradient explosion or disappearance is prevented. The sequence properties of the input are kept since the LSTM may self-transfer and use its own state and output vector in time, such as time t. $\mathbb{C}_{\mathbb{t}}$ defines a kind of storage that can save long-term dependent information input while also collecting historical information states. In the forgetting gate, equation (12) calculates the input data at time t and stores it in the storage cell $\mathbb{C}_{\mathbb{t}-1}$. In the input gate $\mathbb{i}_{\mathbb{t}}$, equation (13) calculates the input data $\mathbb{x}_{\mathbb{t}}$ at time t: the data it determined jointly was collected and $\mathbb{S}_{\mathbb{t}}$ should be updated and placed in the storage cell $\mathbb{C}_{\mathbb{t}-1}$ at time t -1, which becomes $\mathbb{C}_{\mathbb{t}}$ at time t after filtering and updating information. Equation (14) may be used to determine the input data at time t in the output gate: the storage cell $\mathbb{C}_{\mathbb{t}}$outputs information $\mathbb{h}_t$ according to $\mathbb{o}_{\mathbb{t}}$.

$$\mathbb{f}_{\mathbb{t}} = \Sigma(\mathbb{W}_{\mathbb{f}}\mathbb{x}_{\mathbb{t}} + \mathbb{V}_{\mathbb{f}}\mathbb{x}_{\mathbb{t}} + \mathbb{b}_{\mathbb{f}}) \quad (12)$$

$$\mathbb{i}_{\mathbb{t}} = \Sigma(\mathbb{W}_{\mathbb{i}}\mathbb{x}_{\mathbb{t}} + \mathbb{V}_{\mathbb{i}}\mathbb{x}_{\mathbb{t}} + \mathbb{b}_{\mathbb{i}})$$

$$\mathbb{S}_{\mathbb{t}} = \tan\mathbb{h}(\mathbb{W}_{\mathbb{S}}\mathbb{x}_{\mathbb{t}} + \mathbb{V}_{\mathbb{S}}\mathbb{x}_{\mathbb{t}} + \mathbb{b}_{\mathbb{S}}) \quad (13)$$

$$\mathbb{C}_{\mathbb{t}} = \mathbb{f}_{\mathbb{t}} \otimes \mathbb{C}_{\mathbb{t}-1} + \mathbb{i}_{\mathbb{t}} \otimes \mathbb{S}_{\mathbb{t}}$$

$$\mathbb{o}_{\mathbb{t}} = \Sigma(\mathbb{W}_{\mathbb{o}}\mathbb{x}_{\mathbb{t}} + \mathbb{V}_{\mathbb{o}}\mathbb{x}_{\mathbb{t}} + \mathbb{b}_{\mathbb{o}})$$

$$\mathbb{h}_t = \mathbb{o}_{\mathbb{t}} \otimes \tanh(\mathbb{C}_{\mathbb{t}}) \quad (14)$$

where $\mathbb{i}_{\mathbb{t}}$, $\mathbb{f}_{\mathbb{t}}$ and $\mathbb{o}_{\mathbb{t}}$ are input gate, forget gate, and output gate outputs, respectively; $\mathbb{W} \in \mathbb{R}^{d\times k}$ and $\mathbb{V} \in \mathbb{R}^{d\times d}$ are the shared weight matrices, and $\mathbb{b} \in R^d$ is the repeatedly updated shared bias vector; k the size of the concealed vectors is a sign; Σ the activation function is represented by; and $\otimes$ the functioning of the element-wise product.

Transfer learning: Due to its capacity to address issues in connected but unlike domains, voice recognition, image categorization, and medical diagnosis all make considerable use of transfer learning. The transfer learning approach involves transferring the relevant information acquired from the source domain to the target domain, enabling the completion of a defect diagnostic work in the target domain with a limited dataset and various working circumstances. Transfer learning, which uses a large amount of valid label data gathered in the dataset as the auxiliary, may be used to transfer the useful information derived from experimental data to the target task. To address the issue of failure detection techniques brought about by inconsistent data distributions and a lack of reliable data in engineering practice. The use of comparable auxiliary characteristics to solve the target feature is another element of transfer learning. The weight parameters of the model to extract features do not need to be retrained when a new feature is encountered; instead, they may be fine-tuned by entering the weight parameters from the model that was transfer-learned for the source work.

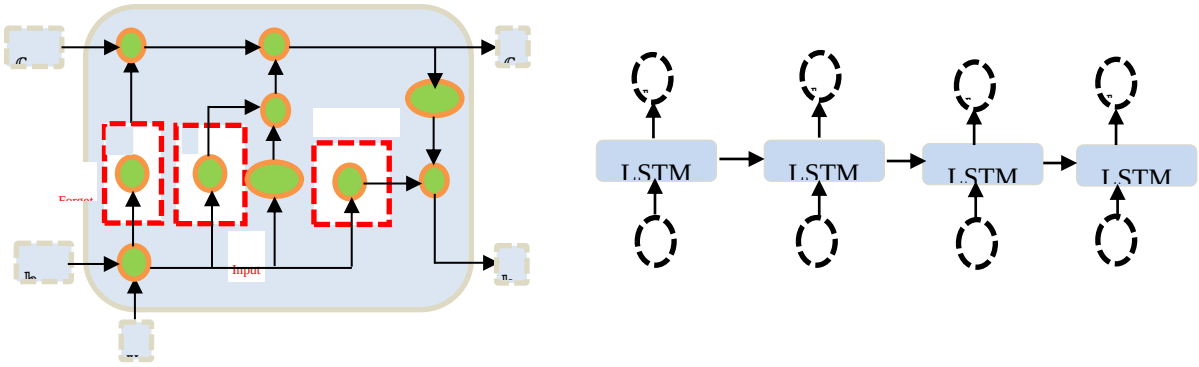

Fig.4. Block diagram of LSTM

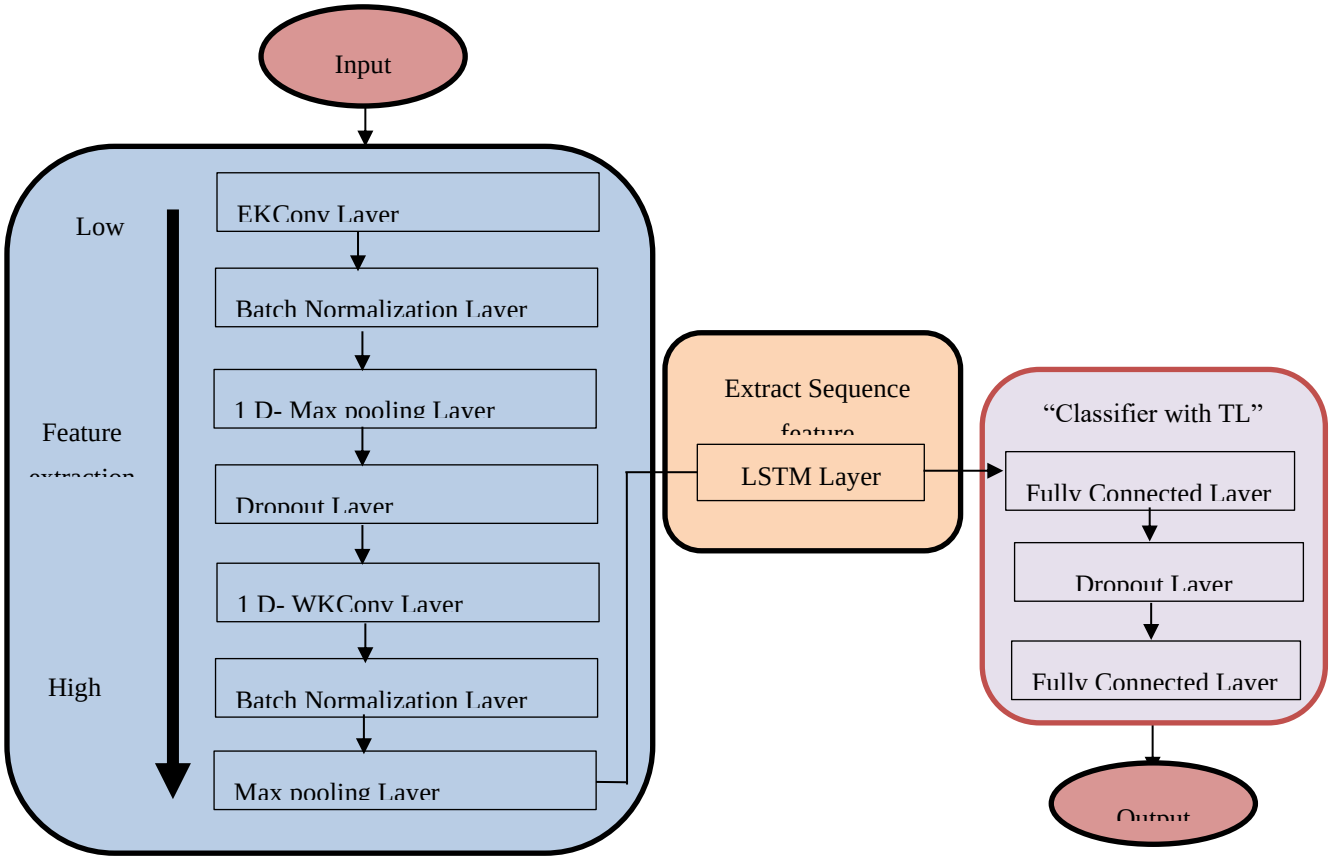

Fig.5. The proposed HEKCNN-LSTM-TL Model

Proposed Method: This paper proposes a transfer learning method based on the HEKCNN-LSTM-TL model that uses one-dimensional attributes from small datasets as input and a fine-tuned transfer learning strategy as shown in Fig. 5 to solve the bearing fault diagnosis problem caused by insufficient data and different data distribution The use of a broad convolution kernel in convolution is acknowledged. This approach allows for the acquisition of a substantial receptive field while minimizing the number of parameters required. Consequently, the model is able to effectively extract low-frequency characteristics from signals.

Additionally, this technique serves to mitigate overfitting of the model, as seen in Figure 5. Data pre-processing, model pre-training, model fine-tuning, and testing make up the four primary steps of the proposed HEKCNN-LSTM-TL technique. A sufficient number of samples are gathered during the training step, as seen in Figure 5, for model pretraining. A limited sample of data is gathered in the testing step to adjust and test the transfer model. To acquire transfer learning weight parameters, the model is pre-trained. Transfer learning is used to transfer relevant model parameters from Ts to Tt [29], and a small amount of target domain training data in Tt is used to fine-tune the model. Using the target domain's test dataset to validate the improved model and provide findings for cancer diagnosis.

4. Experimental Results and discussion

This study employed Tensor-flow open-source deep-learning library as the back-end and Anaconda Python 3.5 as the development environment to build our model. Keras Deep Learning Library served as the front end. The rL-SVM [15], ADBSCAN [113], CNN [18], and HEKCNN-LSTM-TL classifiers were only a few of the classifiers used in this investigation. Hybrid Learning is more accurate and stable than other classifiers. Training and testing groups make up roughly 25% of the datasets. MCRelief-F is used to choose the most pertinent attributes and remove the non-relevant ones. The suggested HEKCNN-LSTM-TL classifier is then used to identify cancer once characteristics are chosen using SOT. A result that appropriately predicts the positive class is referred to as a true positive value (TP) in the context of classification presentation. Another outcome that accurately predicts the negative class is a true negative value (TN). Meanwhile, the misclassified samples are represented by the false negative (FN) and false positive (FP) values. The formulae for accuracy, precision, F1-score, Mathew Correlation-coefficient (MCC), specificity, and sensitivity may be represented as follows using these parameters:

The following formula may be used to compute the accuracy of the classification model, which illustrates the model's overall performance:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100$$

AUC (Area under the ROC Curve). AUC evaluates classification performance across all criteria. The likelihood that a random positive example will be evaluated higher than a random negative example is one way to analyze AUC.

$$AUC = \frac{Recall + sensibility}{2} * 100$$

The weighted indicator of recall accuracy and sensitivity is the F1 score. Between 0 and 1 is its value range. If this parameter's value is 1, the classification algorithm's performance was excellent, and if this parameter's value is 0, the classification algorithm's performance was bad.

$$F1\ score = \frac{2 * Precision * Recall}{Precision + Recall}$$

MCC measures how well the actual and expected outcomes correlate. MCC provides outcome values ranging from -1 to +1. Where 1 denotes the classifier's wholly incorrect guess. The ideal prediction of the classification models is represented by +1, whereas 0 indicates that the classifier creates random predictions. The following formula may be used to determine MCC values:

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

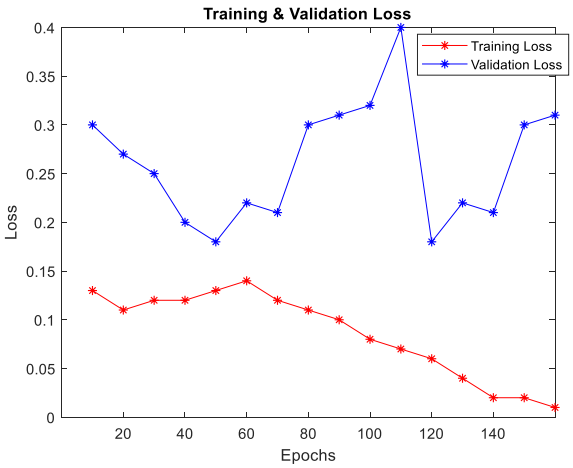

Fig.6. loss vs epoch for proposed learning model

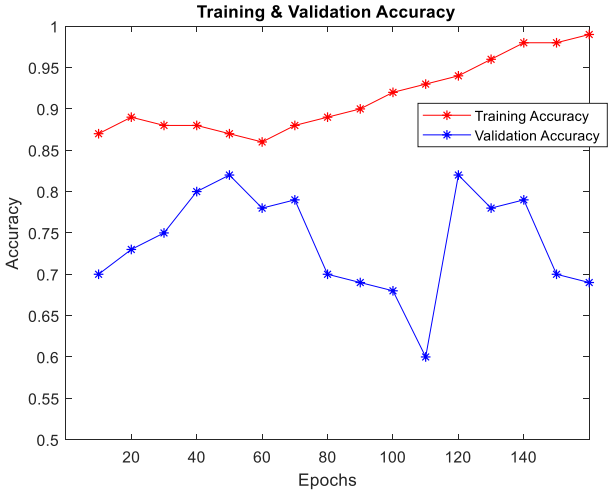

Fig.7. Training accuracy vs epoch for proposed learning model

The ratio of newly identified healthy persons to all healthy people is known as specificity. It indicates that the individual is well and that the forecast was incorrect. The following is the formula for determining specificity:

$$Specificity = \frac{TN}{TN + FP} * 100$$

Sensitivity is the proportion of heart disease patients with new diagnoses to those with all forms of the illness. It implies that the individual has cardiac disease and the model's forecast is favorable. Listed below is the sensitivity calculation formula:

$$Sensitivity = \frac{TP}{TP + FN} * 100$$

Fig.6 and 7 displays the modified HEKCNN-LSTM-TL's training loss and accuracy respectively. Validation loss values for the suggested model have been shown to be lower than those for the baseline model. The first fifty simulation iterations show rapid convergence of the model, but after that, convergence slows down. Starting out, the accuracy was below par, but it has slowly increased as the number of epochs has increased.

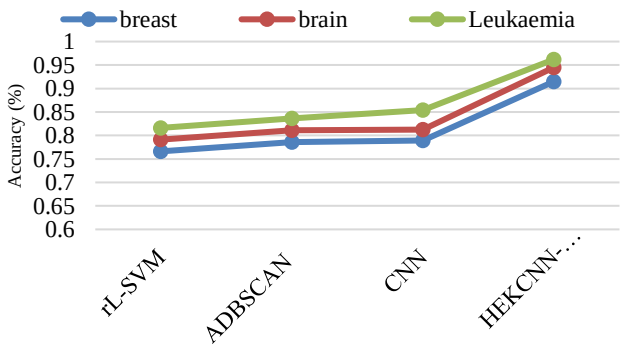

Fig.8. Comparison of the distribution of Accuracy validation values

Table 1. The numerical value of accuracy for proposed and exiting methods

Methods	breast	brain	leukemia
rL-SVM	0.7662	0.7912	0.8162
ADBSCAN	0.7862	0.8112	0.8362
CNN	0.7895	0.8125	0.8542
HEKCNN-LSTM-TL	0.915	0.945	0.962

Therefore, to evaluate the HEKCNN-LSTM-TL method, compared them with existing three ML approaches: rL-SVM, ADBSCAN and CNN and strategies. The four matching Accuracy distributions are shown in Fig. 8. HEKCNN-LSTM-TL performs better than the other techniques when taking into account the average Accuracy values provided. Finally, it should be noted that the TL-based approach produced the highest values for the suggested technique. Since the excellent quality produced by the suggested HEKCNN-LSTM-TL may be used to enhance illness identification, the HEKCNN-LSTM-TL achieves an accuracy of 0.915% (breast), 0.945% (brain), and 0.962% (leukemia) when compared to all other models. Results from current techniques like rL-SVM, ADBSCAN, and CNN are subpar compared to previous techniques. As a consequence, the suggested method has superior validation results for forecasting cancer illness to those of the current algorithms. The numerical value of accuracy for proposed and exiting methods are given in table 1

In Fig. 9, the suggested HEKCNN-LSTM-TL approach's 50 validation AUC value distribution is contrasted with the 50 validation AUC value distributions obtained by the three ML techniques (rL-SVM, ADBSCAN, and CNN). Finally, it's important to note that the TL-based technique yielded the best results for the proposed method. Since the excellent quality obtained by the suggested HEKCNN-LSTM-TL may be employed to improve disease identification, the HEKCNN-LSTM-TL achieves the AUC of 0.88% (breast), 0.89% (brain), and 0.914% (Leukemia) compared to all other models. Results from well-known techniques include rL-SVM (0.7415% for breast cancer, 0.7315% for brain cancer, and 0.7515% for leukemia), ADBSCAN (0.7421% for breast cancer, 0.722% for brain cancer, and 0.7621% for leukemia), and CNN (0.81% for breast cancer, 0.81% for brain cancer, and 0.82% for leukemia).

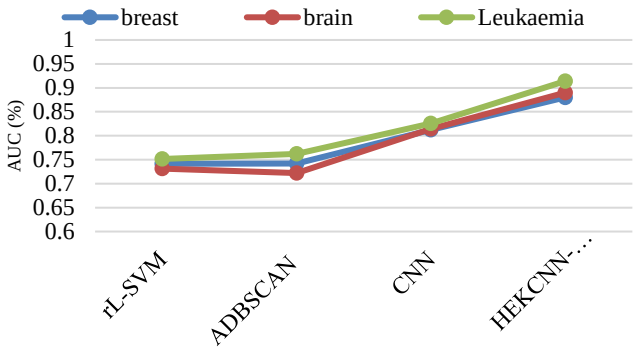

Fig.9. The distribution of AUC validation values

Table 2. The numerical value of AUC for proposed and exiting methods

Methods	breast	brain	leukemia
rL-SVM	0.7415	0.7315	0.7515
ADBSCAN	0.7421	0.7221	0.7621
CNN	0.8125	0.8145	0.8254
HEKCNN-LSTM-TL	0.88	0.89	0.914

As a result, in terms of robust validation results, the suggested technique performs better than the state-of-the-art cancer sickness prediction algorithms. Existing approaches fail to provide good detection outcomes when the features are denoised. The suggested method outperforms the other three in its ability to overcome the influence of pre-processing features and detect malignancy. The numerical value of AUC for proposed and exiting methods are given in table 2.

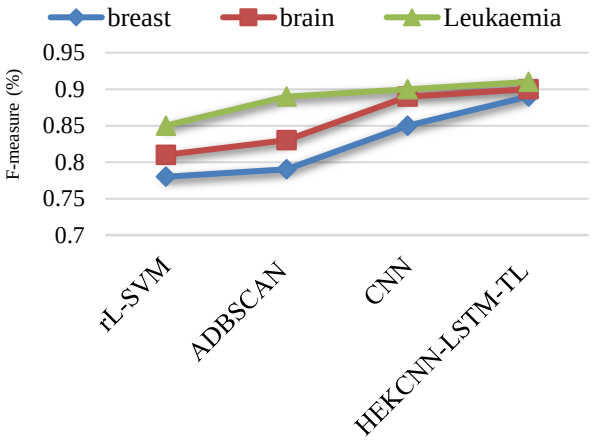

Fig.10. The distribution of F-Measure validation values

The HEKCNN-LSTM-TL outperforms other methods including rL-SVM, ADBSCAN, and CNN, as seen in Fig. 10. Last but not least, it should be noted that the TL-based approach produced the greatest results for the suggested

procedure. In comparison to all other models, the HEKCNN-LSTM-TL achieves the f-measure of 0.89% (breast), 0.90% (brain), and 0.91% (leukemia) due to the excellent quality that is created by the suggested HEKCNN-LSTM-TL. According to current approaches like rL-SVM, ADBSCAN, and CNN, the findings are (0.78% breast, 0.81% brain, and 0.85% leukemia), (0.79% breast, 0.83% brain, and 0.89% leukemia), and (0.85% breast, 0.89% brain, and 0.90% leukemia), respectively. As a consequence, the suggested method has superior validation results for classifying cancer illness to those of the current algorithms. The CNN approach produces the poorest results for cancer illness identification because it is sensitive to weak edges. The ADBSCAN approach often detects the area of cancer illness that is the brightest, which yields incorrect detection findings. The suggested strategy, however, makes use of the deep characteristics that are derived from the pre-trained HEKCNN-LSTM-TL is less affected by the noise cancer disease data and the optimal features obtains the desirable cancer diseases detection results. The numerical value of f-measure for proposed and exiting methods are given in table 3.

Table 3. The numerical value of f-measure for proposed and exiting methods

Methods	breast	brain	leukemia
rL-SVM	0.78	0.81	0.85
ADBSCAN	0.79	0.83	0.89
CNN	0.85	0.89	0.9
HEKCNN-LSTM-TL	0.89	0.9	0.91

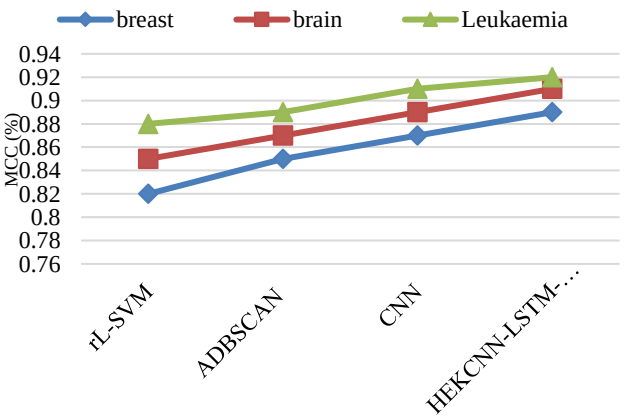

Fig.11. The distribution of MCC validation values

The four corresponding MCC distributions are shown in Fig. 11. HEKCNN-LSTM-TL performs better than the other techniques when taking into account the average Accuracy values provided. Finally, it should be noted that the TL-based approach produced the highest values for the suggested technique. The suggested HEKCNN-LSTM-TL outperforms all previous models in terms of MCC, achieving values of 0.89% (breast), 0.91% (brain), and 0.92% (leukemia), due to the excellent quality data it generates that may be leveraged to enhance illness identification. Thus, the suggested approach improves validation findings for cancer prediction. The effectiveness of current techniques may be determined by analyzing the experimental data as follows: In comparison to the approaches previously stated, the suggested method is more effective at extracting the borders of the cancer illness via efficient clustering and may, to a certain degree, resist the effects of lesion interference. The numerical value of MCC for proposed and exiting methods are given in table 4.

Table 4. The numerical value of MCC for proposed and exiting methods

Methods	breast	brain	leukemia
rL-SVM	0.82	0.85	0.88
ADBSCAN	0.85	0.87	0.89
CNN	0.87	0.89	0.91
HEKCNN-LSTM-TL	0.89	0.91	0.92

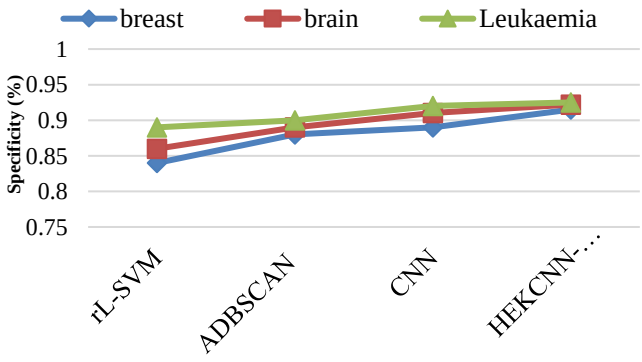

Fig.12. The distribution of Specificity validation values

Since the excellent quality produced by the suggested HEKCNN-LSTM-TL may be used to enhance illness diagnosis, the graph in Figure 12 demonstrates that it achieves specificity of 0.915% (breast), 0.922% (brain), and 0.925% (leukemia) compared to all other models. The results of existing methods such as rL-SVM (0.84%-breast, 0.86%-brain and 0.89%- Leukemia), ADBSCAN (0.88%-breast, 0.89%-brain and 0.9%- Leukemia) and CNN

(0.89%-breast, 0.91%-brain and 0.92%- Leukemia) respectively. As a consequence, the suggested algorithm performs better in terms of validation findings for predicting cancer illness than the current algorithms. The findings show that the suggested technique has produced results that are comparable to those of the other methods and is superior since it can identify previous cancer illness using an effective EFFCM and adaptive Otsu threshold with the ALO segmentation method and, if required, adjust the graph as efficiently. The numerical value of specificity for proposed and exiting methods are given in table 5.

Table 5. The numerical value of specificity for proposed and exiting methods

Methods	breast	brain	leukemia
rL-SVM	0.84	0.86	0.89
ADBSCAN	0.88	0.89	0.9
CNN	0.89	0.91	0.92
HEKCNN-LSTM-TL	0.915	0.922	0.925

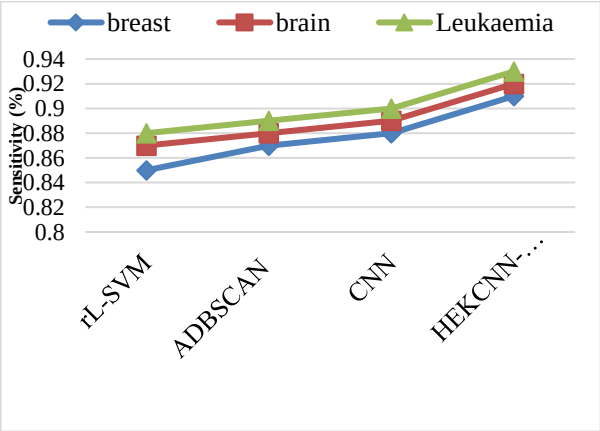

Fig.13. The distribution of Sensitivity validation values

In Fig.13, graph shows the HEKCNN-LSTM-TL attains the accuracy of 0.91% (breast), 0.92% (brain) and 0.93% (Leukemia) HEKCNN-LSTM-TL has been offered as a superior model to all others since it can be used to create data of excellent quality that will help with illness identification. Therefore, in terms of improved validation findings for predicting cancer illness, the suggested algorithm is superior to the current methods. As a result of this suggested method's propensity to emphasize detection efficiency while simultaneously taking into account target detection accuracy, it also benefits from the feature selection module. The numerical value of accuracy for proposed and exiting methods are given in table 6.

Table 6. The numerical value of accuracy for proposed and exiting methods

Methods	breast	brain	leukemia
rL-SVM	0.85	0.87	0.88
ADBSCAN	0.87	0.88	0.89
CNN	0.88	0.89	0.9
HEKCNN-LSTM-TL	0.91	0.92	0.93

Fig. 14-16 depicts that the value of fitness vs number of iterations for three datasets such as brain, breast and leukaemia correspondingly. At the 70th iteration, the performance of rL-SVM, ADBSCAN, CNN, and HECKCNN-LSTM-TL is assessed. In contrast, this iteration yields 16.56, a global optimum route, indicating that GA still needs time to converge. 20 iterations identify HECKCNN-LSTM-TL's optimal route. Table 6 shows the iteration route and fitness value.

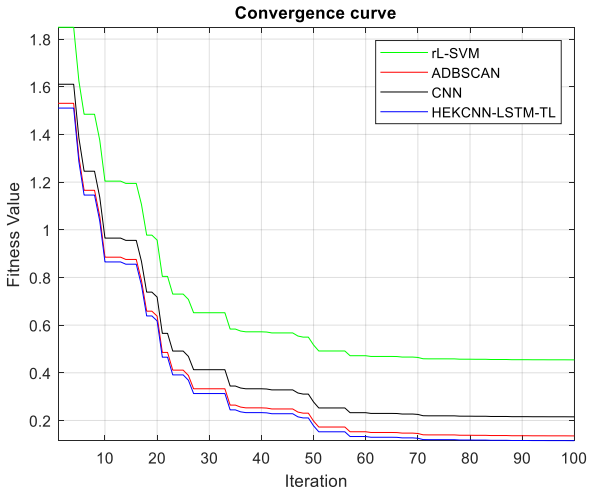

Fig.14. Brain dataset convergence graph

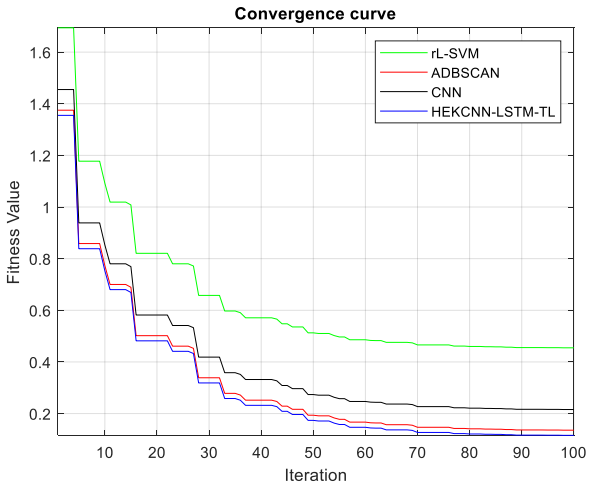

Fig.15. Breast dataset convergence graph

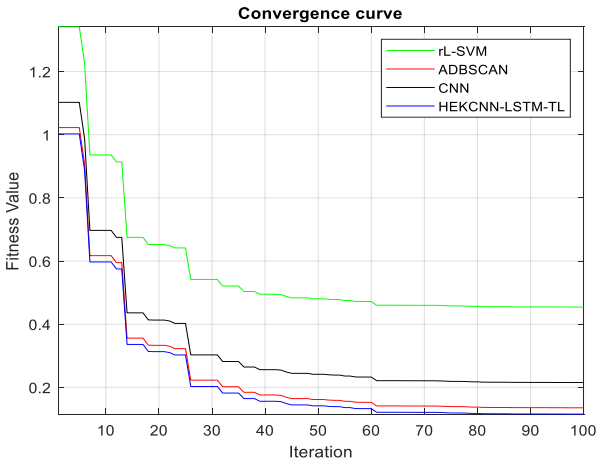

Fig.16. Leukemia dataset convergence graph

The convergence curve in Fig. 14–16, which reveals that the HEKCNN-LSTM–TL method is the first algorithm to reach the greatest accuracy for the input dataset, demonstrates the proposed model's successful performance in this test. The SOT method with differential operator is suggested to address the following issues and tested for optimal fitness, average number of function evaluations, and convergence rate. The SOT-based algorithm outperforms existing nature-inspired optimization approaches for feature selection, according to the findings.

5. Conclusion and Future work

HEKCNN-LSTM-TL is a novel technique for cancer prediction utilizing gene expression data. The suggested methodology integrates the efficiency of an EKCNN model-LSTM for extracting significant features from structured input data, along with the effectiveness of a transfer learning (TL) approach to address the problem of overfitting in situations involving limited training datasets containing high-dimensional samples. This combination is particularly relevant in the context of cancer prediction tasks utilizing gene-expression data, which often exhibit these characteristics. SOT feature selection has been presented as a way to enhance the performance of cancer survival prediction since the initial RNA-Seq samples lack the local structure required for MCRelief-F filters to function successfully as feature extractors. In order to do this, gene-expression vectors are presented onto images by applying the well-known Heatmap approach and adhering to domain-specific criteria for this rearrangement, which uses the Kaggle database to query the genes in the dataset. A reduced dataset made up of hundreds of tumor-labeled samples was used to refine the pre-trained HEKCNN-LSTM-TL. In order to assess the effectiveness of the HEKCNN-LSTM-TL approach, we have contrasted its outcomes with those obtained by a variety of more traditional ML techniques. These latter methods combine conventional ML models with various dimensionality reduction techniques. The AUC shows that the HEKCNN-LSTM-TL strategy is superior to the other ML approaches considered here. In instance, CNN achieved the highest performance rate. The implications of these finds are significant in terms of the early detection of cancer of the stomach, as well as the development of therapeutic approaches and prognostic tools. These advancements are based on the examination of candidate genes and their expression patterns, using deep learning techniques and large-scale data analysis. Using a parallel computing strategy helps lessen the load on the computer during the classification of these data sets. There will come a time when many multi-objective optimization techniques can be used to streamline the process of feature selection and categorization. In addition, future research should make use of the newest gene expression dataset or expand beyond microarray datasets to include other types of datasets, such as RNA-Sequence.

References

[1] Sung H., Ferlay J., Siegel R. L., Laversanne M., Soerjomataram I., Jemal A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 71 (3), 209–249.

[2] Id J. L., Zhou Z., Dong J., Fu Y., Li Y., Luan Z., et al. (2021). Predicting breast cancer 5-year survival using machine learning: A systematic review. *PLoS One* 16, e0250370–23.

[3] Chandra, B., & Gupta, M. (2011). An efficient statistical feature selection approach for classification of gene expression data. *Journal of biomedical informatics*, *44*(4), 529-535.

[4] Huang, S., Yang, J., Fong, S., & Zhao, Q. (2020). Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges. *Cancer letters*, *471*, 61-71.

[5] Daoud, M., & Mayo, M. (2019). A survey of neural network-based cancer prediction models from microarray data. *Artificial intelligence in medicine*, *97*, 204-214.

[6] Liu, B., Cui, Q., Jiang, T., & Ma, S. (2004). A combinational feature selection and ensemble neural network method for classification of gene expression data. *BMC bioinformatics*, *5*, 1-12.

[7] Shah, S. H., Iqbal, M. J., Ahmad, I., Khan, S., & Rodrigues, J. J. (2020). Optimized gene selection and classification of cancer from microarray gene expression data using deep learning. *Neural Computing and Applications*, 1-12.

[8] Manikandan, G., & Abirami, S. (2018). A survey on feature selection and extraction techniques for high-

dimensional microarray datasets. *Knowledge Computing and its Applications: Knowledge Computing in Specific Domains: Volume II*, 311-333.

[9] Musheer, R. A., Verma, C. K., & Srivastava, N. (2019). Novel machine learning approach for classification of high-dimensional microarray data. *Soft Computing*, *23*, 13409-13421.

[10] Daoud, M., & Mayo, M. (2019). A survey of neural network-based cancer prediction models from microarray data. *Artificial intelligence in medicine*, *97*, 204-214.

[11] Gupta, S., Gupta, M. K., Shabaz, M., & Sharma, A. (2022). Deep learning techniques for cancer classification using microarray gene expression data. *Frontiers in Physiology*, 2022.

[12] Kar, S., Sharma, K. D., & Maitra, M. (2015). Gene selection from microarray gene expression data for classification of cancer subgroups employing PSO and adaptive K-nearest neighborhood technique. *Expert Systems with Applications*, *42*(1), 612-627.

[13] Mishra, P., & Bhoi, N. (2021). Cancer gene recognition from microarray data with manta ray based enhanced ANFIS technique. *Biocybernetics and Biomedical Engineering*, *41*(3), 916-932.

[14] Halder, A., & Kumar, A. (2019). Active learning using rough fuzzy classifier for cancer prediction from microarray gene expression data. *Journal of Biomedical Informatics*, *92*, 103136.

[15] Mazlan, A. U., Sahabudin, N. A., Remli, M. A., Ismail, N. S. N., & Adenuga, K. I. (2021). An enhanced feature selection and cancer classification for microarray data using relaxed Lasso and support vector machine. In *Translational Bioinformatics in Healthcare and Medicine* (pp. 193-200). Academic Press.

[16] Shafi, A. S. M., Molla, M. I., Jui, J. J., & Rahman, M. M. (2020). Detection of colon cancer based on microarray dataset using machine learning as a feature selection and classification techniques. *SN Applied Sciences*, *2*, 1-8.

[17] Hambali, M. A., Oladele, T. O., Adewole, K. S., Sangaiah, A. K., & Gao, W. (2022). Feature selection and computational optimization in high-dimensional microarray cancer datasets via InfoGain-modified bat algorithm. *Multimedia Tools and Applications*, *81*(25), 36505-36549.

[18] Alrefai, N., Ibrahim, O., Shehzad, H. M. F., Altigani, A., Abu-ulbeh, W., Alzaqebah, M., & Alsmadi, M. K. (2022). An integrated framework based deep learning for cancer classification using microarray datasets. *Journal of Ambient Intelligence and Humanized Computing*, 1-12.

[19] Alrefai, N., & Ibrahim, O. (2022). Optimized feature selection method using particle swarm intelligence with ensemble learning for cancer classification based on microarray datasets. *Neural Computing and Applications*, *34*(16), 13513-13528.

[20] Khaire, U. M., & Dhanalakshmi, R. (2020). High-dimensional microarray dataset classification using an improved adam optimizer (iAdam). *Journal of Ambient Intelligence and Humanized Computing*, *11*(11), 5187-5204.

[21] Zhang, D., Zou, L., Zhou, X., & He, F. (2018). Integrating feature selection and feature extraction methods with deep learning to predict clinical outcome of breast cancer. Ieee Access, 6, 28936-28944.

[22] Brain cancer gene expression - CuMiDa | Kaggle

[23] Breast cancer gene expression - CuMiDa | Kaggle

[24] Leukemia gene expression - CuMiDa | Kaggle

[25] Lattimore, T. (2016). Regret analysis of the anytime optimally confident UCB algorithm. *arXiv preprint arXiv:1603.08661*.

[26] Khare, A., Kakandikar, G. M., & Kulkarni, O. K. (2022). An Insight Review on Jellyfish Optimization Algorithm and Its Application in Engineering. *Journal homepage: http://iieta. org/journals/rces*, *9*(1), 31-40.

[27] Zeebaree, D. Q., Haron, H., & Abdulazeez, A. M. (2018, October). Gene selection and classification of microarray data using convolutional neural network. In *2018 International Conference on Advanced Science and Engineering (ICOASE)* (pp. 145-150). IEEE.

[28] Şahín, C. B., & Diri, B. (2019). Robust feature selection with LSTM recurrent neural networks for artificial immune recognition system. *IEEE Access*, *7*, 24165-24178.

[29] George, B., Gokhale, S. D., Yaswanth, P. M., Vijayan, A., Devika, S., & Suchithra, T. V. (2022). Identification of Alzheimer associated differentially expressed gene through microarray data and transfer learning-based image analysis. *Neuroscience Letters*, *766*, 136357.

[30] Chou, J. S., & Molla, A. (2022). Recent advances in use of bio-inspired jellyfish search algorithm for solving optimization problems. *Scientific*

Reports, *12*(1), 19157.

[31] Rani, R. U. ., Rao, P. S. ., Lavanaya, K. ., Satyanarayana, N. ., Lallitha, S. ., & Prasad J, P. . (2023). Optimization of Energy-Efficient Cluster Head Selection Algorithm for Internet of Things in Wireless Sensor Networks. International Journal on Recent and Innovation Trends in Computing and Communication, 11(4), 238–248. https://doi.org/10.17762/ijritcc.v11i4.6445

[32] Smith, J., Ivanov, G., Petrović, M., Silva, J., & García, A. Detecting Fake News: A Machine Learning Approach. Kuwait Journal of Machine Learning, 1(3). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/142

[33] Dhabliya, D. (2019). Security analysis of password schemes using virtual environment. International Journal of Advanced Science and Technology, 28(20), 1334-1339. Retrieved from www.scopus.com