IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 14 November 2023, accepted 19 December 2023, date of publication 25 December 2023,
date of current version 5 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3346925

RESEARCH ARTICLE

Long Short-Term Memory-Deep Belief Network-Based Gene Expression Data Analysis for Prostate Cancer Detection and Classification

BIJAYA KUMAR SETHI[1], DEBABRATA SINGH[2], SAROJA KUMAR ROUT[3], (Senior Member, IEEE), AND SANDEEP KUMAR PANDA[4], (Member, IEEE)

[1]Department of Computer Science and Engineering, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha 751030, India
[2]Department of Computer Application, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha 751030, India
[3]Department of Information Technology, Vardhaman College of Engineering (Autonomous), Hyderabad, Telangana 501218, India
[4]Department of Artificial Intelligence and Data Science, Faculty of Science and Technology (IcfaiTech), ICFAI Foundation for Higher Education, Hyderabad, Telangana 501203, India

Corresponding author: Sandeep Kumar Panda (sandeeppanda@ifheindia.org)

This work was supported in part by LIVIA Laboratory, Faculty of Science and Technology (IcfaiTech) and in part by the ICFAI Foundation for Higher Education, Hyderabad, Telangana, India.

ABSTRACT Prostate cancer (PRC) is the major reason of mortality globally. Early recognition and classification of PRC become essential to enhance the quality of healthcare services. A newly established deep learning (DL) and machine learning (ML) approach with different optimization tools can be employed to classify accurately of PRC accurately using microarray gene expression data (GED). Though the microarray data structures are important to diagnosing different kinds of diseases, the optimum hyperparameter tuning of the DL models poses a major challenge to achieving maximum classification performance. To resolve these issues, this study develops a new Gene Expression Data Analysis using Artificial Intelligence for Prostate Cancer Diagnoses (GEDAAI-PCD) technique. The proposed GEDAAI-PCD technique examines the GED for the identification of PRC. To accomplish this, the GEDAAI-PCD technique initially normalizes the GED into a uniform format. In addition, the long short-term memory-deep belief network (LSTM-DBN) model was applied for PRC classification purposes. The wild horse optimization (EWHO) system was utilized as a hyperparameter tuning strategy to optimize the performance of the LSTM-DBN model. The experimental assessment of the GEDAAI-PCD system occurs on open open-accessed gene expression database. The experimental outcomes emphasized the supremacy of the GEDAAI-PCD method on PRC classification.

INDEX TERMS Prostate cancer, deep learning, artificial intelligence, microarray gene expression, parameter tuning.

I. INTRODUCTION

The introduction of the research paper titled Long Short Term Memory-Deep Belief Network based Gene Expression Data Analysis for Prostate Cancer Detection and Classification is a pivotal section that offers a comprehensive overview of the study's scope, significance, and objectives. Prostate cancer stands as one of the most widespread malignancies affecting men on a global scale. Timely and precise detection, as well as accurate classification of prostate cancer, are crucial factors contributing to effective interventions and ultimately enhanced patient outcomes. Conventional techniques for diagnosing prostate cancer frequently necessitate labor-intensive manual feature engineering procedures, which often fall short in capturing subtle intricacies inherent in gene expression data. In light of these challenges, this study ventures into uncharted territory by embarking on an innovative approach that fuses the strengths of Long Short-Term Memory (LSTM) and Deep Belief Network (DBN) methodologies for the analysis of gene expression data. LSTM

The associate editor coordinating the review of this manuscript and approving it for publication was Wai-Keung Fung.

emerges as the optimal choice for modeling sequential data, thereby offering an ideal framework for capturing temporal dependencies within gene expression profiles. Conversely, DBN exhibits remarkable prowess in the extraction of hierarchical features from multifaceted data structures. The second most common reason for death in men is prostate cancer (PRC). It derives afterward lung cancer concerning threatening and exceeding in the domain. The etiology of this kind of cancer is still not completely identified yet [1]. However a few features like environmental, heredity, and diet controls that affect male hormone have already been discussed in epidemiological research. Microarray technology is a new technology from molecular biology with regards to the contribution of data in quantifying millions of genes that are used during the diagnoses of disease and to forecast the potential outcome [2]. The gene that is controlled owing to disease conditions could be examined by the expression extracting in the microarray information [3]. As well, this measurement assists in the examination of cancer. The usage of the DNA microarray technique to expose data from the expression level of millions of genes has more potential [4]. DNA microarray technology could define the level of millions of genes concurrently in a single experiment. The study of gene expression is crucial in different subject areas of biological analysis to attain essential data.

On the other hand, Machine learning (ML) techniques were effectively employed on PRC information to recognize gene biomarkers of diseases [5]. Automating the separation of genes in microarray information, can decrease the classification error and decrease the time factor included while completing the process. A feature selection in ML intends to achieve the least subset of problem space although quite accomplishing the maximum level of detection and classification [6]. Likewise, feature selection including embedded procedures, filter models, and wrapper models, an optimization strategy could provide satisfactory outcomes [7]. Filtering and wrapping are well-known techniques for the selection of genes. According to the filtering method, all the features are allocated a value according to their relationship with a single variable scoring criterion and a class label [8]. Subsequently, the gene with the maximum ranking is selected and categorized. At the same time, the wrapper technique requires a set of classifiers to evaluate every gene efficiency from the ranking procedure [9]. Therefore, the optimum subclass of the gene can be defined according to the performance scoring or rankings in every discovered subcategory. Even though the filtering method could not measure genomic relationship, the wrapper strategy might be constrained by their highest processing cost. Using data from an ICU, predict whether a patient has Sepsis Disease with a Deep Neural Network. [10]. The authors represent a hybrid evolution-based deep learning model that takes advantage of multimodal data. In a multi-modal fusion framework, histopathological images are combined with gene modality. Each modality's state and form were taken into account to build the deep feature extraction network [11]. High-performance computing and machine learning algorithms are combined for better medical image analysis techniques like fused, segmented, registered, and classified images [12]. Deep learning algorithms, unlike classical neural networks, do not depend on feature extraction. The medical field has a growing need for computer-aided automatic processing due to its enhanced accuracy and precision [12]. The established multi-modal fusion framework combines gene modality with histopathological image modality. We establish separate deep feature extraction networks for each modality, taking into account their distinct states and forms [11]. The machine learning model was subsequently customized to forecast the D'Amico Risk Classification. It was trained using information from a cohort of 54 prostate cancer patients, demonstrating its capability to effectively distinguish between low-/intermediate-risk and high-risk diseases, all without necessitating supplementary clinical data [13]. We conducted a prospective study wherein we performed biopsies on a cohort of men exhibiting benign findings during digital rectal examinations (DRE), with a PSA level below 20ngml−1, and no prior history of prostate biopsies [14].

The research gap in the study Long Short-Term Memory-Deep Belief Network based Gene Expression Data Analysis for Prostate Cancer Detection and Classification. The field of gene expression data analysis is evolving rapidly, yet the underutilization of deep learning methods, like Long Short-Term Memory-Deep Belief Networks (LSTM-DBN), in prostate cancer detection and classification presents a significant research gap. The combination of Long Short-Term Memory (LSTM) and Deep Belief Networks (DBN) for gene expression analysis in prostate cancer is relatively unexplored, creating a gap in understanding the effectiveness of such hybrid models. While LSTM-DBN models may offer a promising approach, the lack of comprehensive performance evaluation, including comparisons with existing methods, hinders their validation. Assessing metrics such as accuracy, sensitivity, and specificity is essential to ascertain their strengths and weaknesses. Insufficient access to diverse and representative datasets for prostate cancer gene expression analysis presents another research gap. Larger, more varied datasets are needed to capture the full biological spectrum of the disease. Moreover, the ethical and clinical implications of applying advanced AI techniques in healthcare, including addressing data biases, privacy concerns, and seamless clinical integration, have not received adequate attention, warranting in-depth exploration. Additionally, the computational intensity of deep learning models, especially LSTM-DBN hybrids, poses challenges in scalability and practical deployment, further contributing to the research gap.

This research endeavors to pioneer an inventive approach for the detection and classification of prostate cancer, harnessing advanced deep learning techniques like Long Short-Term Memory (LSTM) and Deep Belief Networks (DBN). The primary focus is on the analysis of gene expression

data associated with prostate cancer, intending to enhance the precision and dependability of diagnostic processes through the use of hybrid LSTM-DBN models. Key objectives encompass evaluating model efficacy, conducting performance assessments, and comparative analyses with existing methods. The research also underscores the imperative for more extensive and representative datasets, considers ethical aspects in healthcare integration, explores scalability challenges, and seeks to improve the interpretability of AI models. The ultimate aspiration is to propel the field of precision medicine in prostate cancer care through the utilization of LSTM-DBN-based gene expression analysis.

A. MOTIVATION

The impetus for researching Long Short Term Memory-Deep Belief Network based Gene Expression Data Analysis for Prostate Cancer Detection and Classification is underpinned by several compelling factors.

- The complexity and dynamic nature of gene expression patterns is the main driving force behind the use of an LSTM-DBN for gene expression data analysis in the diagnosis and classification of prostate cancer. Accurate detection and classification of prostate cancer samples necessitates sophisticated approaches due to the dynamic and variable nature of gene expression profiles. This can be very important in gene expression research to find complex patterns that could represent various cancer stages. For processing such high-dimensional data and extracting meaningful representations, LSTMs and DBNs are excellent choices.
- Prostate cancer, a global health concern for men, necessitates early identification to improve outcomes. Traditional diagnostic methods involve time-consuming manual processes and struggle to detect subtle gene expression patterns.
- Deep learning and artificial intelligence advancements have transformed various industries, particularly healthcare. They excel in image analysis and language processing. Applying deep learning to gene expression analysis improves cancer detection reliability.
- Data-driven medicine emphasizes using extensive genetic data for informed disease diagnosis and treatment decisions. Analyzing gene expression data provides insights into fundamental molecular mechanisms, benefiting diseases like prostate cancer.
- Integrating AI models into clinical workflows enhances diagnosis accuracy and decision-making in healthcare. This research aims to bridge the gap between advanced machine learning and practical healthcare applications.
- Rising AI impact in healthcare necessitates addressing ethical concerns like data privacy, bias, and transparency. The research aims to establish responsible AI frameworks and healthcare-specific guidelines.
- The use of LSTM-DBN in the analysis of gene expression data related to prostate cancer is important because it can manage the intricacies of high-dimensional, temporal data, improving accuracy and yielding insights that can be useful for both diagnosis and therapy planning. The temporal dependencies in gene expression data can be more accurately modeled with the use of LSTM-DBN. This may result in the identification and categorization of various prostate cancer stages and subtypes with greater accuracy. The deep learning architecture facilitates the automatic extraction of pertinent information from the unprocessed gene expression data. This is important to recognize small alterations and trends that may indicate certain cancer stages. Since LSTMs are meant to mimic sequences, they are appropriate for gene expression data that changes over.
- The capacity to model dynamically is crucial for comprehending how prostate cancer progresses and how gene expression varies. This method helps to create more individualized and focused treatment plans by correctly identifying various prostate cancer subtypes based on gene expression profiles. It may be possible to identify changes in gene expression patterns before they appear clinically thanks to LSTMs' capacity to record temporal relationships. For cancer to be effectively treated, early detection is essential.

B. CONTRIBUTIONS

- A novel technique, Gene Expression Data Analysis using Artificial Intelligence for Prostate Cancer Diagnoses (GEDAAI-PCD), analyzes gene expression data (GED) to identify prostate cancer. It begins by normalizing GED.
- LSTM addresses gradient vanishing in RNNs, boasting high learning capacity with dropout layers. ResNet-101 CNN, with convolution, inception, and fully connected layers, trained on ImageNet, outperforms LSTM in detection.
- LSTM-DBN applied to gene expression data for prostate cancer detection is a promising use of deep learning, offering more accurate diagnoses and advancing our understanding of disease mechanisms and treatment.
- LSTM-DBN is used for prostate cancer classification, employing enhanced wild horse optimization (EWHO) for performance enhancement. The GEDAAI-PCD algorithm is assessed on an open-access gene expression database. The study evaluates LSTM-DBN's efficacy in prostate cancer analysis.
- A thorough performance assessment will use key metrics (accuracy, sensitivity, specificity, F1-score) to measure the LSTM-DBN model's ability to distinguish between prostate cancer and non-cancer cases, emphasizing its role in early detection.
- The study explores using model insights for tailored treatment strategies, aligning with the broader aim of advancing precision medicine in prostate cancer care.

The primary research inquiry in Long Short-Term Memory-Deep Belief Network based Gene Expression Data Analysis

for Prostate Cancer Detection and Classification likely centers around the following research question:

Q1. How can we harness Long Short-Term Memory-Deep Belief Network models to enhance the precision and efficiency of gene expression data analysis for prostate cancer detection and classification.

Q2. What are the wider consequences of this methodology for the field of medical diagnostics and the quality of patient care?

II. RELATED WORKS

An integrated approach for gene expression in disease types can be explained in [15], containing two stages: recognizing the effectual genes using soft ensemble and categorizing them with new DNN. This research work examines in detail the theory of deep learning for the detection of medical anomalies [16]. To develop a bidirectional Long Short-Term Memory Deep Neural Network (biLSTM) model that detects prostate cancer in men based on previously established phenotypic features [17]. To assess prediction performance, the study utilized The Cancer Genome Atlas (TCGA) database, evaluating six GS methods and seven omics data combinations. The findings indicate that the Best Linear Unbiased Prediction (BLUP) model outperforms other methods in terms of predictability and computational efficiency.

Our findings reveal that combining a SIG-HES6 signature with DESNT significantly enhances the prediction of poor outcomes in prostate cancer. We propose a model suggesting cooperation between the SIG-HES6 and SIG-DESNT pathways, which has implications for therapeutic design [18], [19].

The feature selection (FS) scheme integrates 3 approaches for selecting wrapper genes and ranks them based on the KNN technique, which leads to a very generalizable method with a lower error level. Utilizing soft ensembling, the most effectual subsets of genes can be recognized in 3 microarray databases of diffuse huge cell lymphoma, PRC, and leukemia. Gumaei et al. [20] examine utilizing a correlation FS (CFS) approach with random committee (RC) ensemble learning for detecting PRC in microarray data of gene expressions. A series of experimentations can demonstrate that an open benchmark database utilizes a 10-fold cross-validation system to evaluate the projected system. Using an open-source HGSOC proteomic dataset, we used a Machine Learning-based pipeline to create a decision support system (DSS) that was capable of discriminating between HGSOC biopsies [21]. An ensemble learning method is proposed for classifying cancer in the present study. Particle swarm optimization and ensemble learning are used in this publication for feature selection and cancer classification [22].

The authors in [23] established P-NET as a physically learned DL approach for the stratification of the patients with PRC by treatment-resistance state and estimated molecular driver of treatment resistances for therapeutic targeting with interpretability. In this paper, we present an approach to detect ovarian cancer that uses simultaneous feature weighting and parameter optimization. With adaptive differential evolution (ADE) as a fitness function, the weights are optimized with cross-validation errors, least absolute shrinkage, and selection operator regularization [24]. In this study, the author developed a tool for analyzing gynecological ultrasound data using a machine-learning algorithm to predict 12-month PFS in patients with OC [25], [26].

It is illustrated that P-NET forecasts cancer state employing molecular data with effectiveness is higher than other modeling systems. In [10], an integration of effectual ML techniques for FS and classifier can be presented for analyzing gene activities and choosing them as appropriate biomarkers for distinct laterality instances. The projected technique allows to identification of groups of genes that recognize distinct laterality class labels. The resulting genes can be established that strongly correlate with disease progression. A radiomics model that can forecast CCR5 expression levels was developed by utilizing the Cancer Imaging Archive (TCIA) and Cancer Genome Atlas (TCGA) databases to determine the predictive significance of C-C motif chemokine receptor type 5 (CCR5) expression levels in ovarian cancer patients [27]. This suggested model can be used by physicians to detect ovarian cancer much more accurately, increasing the chances of effective treatment and a longer life expectancy for patients [28].

In [29] the authors propose a new AIFSDL-PCD detection method based on the combined capabilities of AIFSDL and DL. Moreover, an FS system based on chaotic invasive weed optimization (CIWO) to select a high-grade feature set illustrates the innovative nature of this work. Furthermore, the DNN technique was executed as a classifier method for detecting the presence of PRCs from the microarray GED. Pandit et al. [30] presented an effectual and hybrid DL approach for molecular cancer classifiers utilizing expression data to solve these restrictions. Afterward, the clustering was completed with the use of an enhanced binomial clustering system. Afterward, the data was extracted with the utilization of a multi-fractal Brownian motion system. Next, essential features can be chosen with the utilization of an enhanced cuckoo search-optimized system. Eventually, the data classifier was carried out utilizing a wavelet-based deep CNN. The authors in [31] established an infrastructure called GraphChrom for cancer classifiers. Chromosomal aberrations (CA) are addressed with GraphChrom, a network neural network (NN) that generates local connectivity between aberrations. In this process, fuzzy c-means and k-means clustering are applied to preprocessed images with better quality [37].

This article introduces a feature selection method for prostate cancer detection, termed Artificial Intelligence-based Feature Selection with Deep Learning for Prostate Cancer Detection (AIFSDL-PCD), utilizing microarray gene expression data. The AIFSDL-PCD technique incorporates pre-processing steps aimed at improving the quality of the input data. The author conducts comprehensive experiments on eight benchmark high-dimensional gene expression datasets, comparing the proposed approach with other

contemporary techniques. Three classifiers, specifically support vector machines (SVM), Naive Bayes (NB), and K-nearest neighbors (KNN), are utilized to evaluate the effectiveness of the selected genes and their influence on classification accuracy. The paper concludes by delving into a discourse on the future of machine learning, contemplating potential advancements in the field that might lead to the development of innovative systems. The objective of the presented research is to determine the most effective strategy utilizing state-of-the-art computer vision, specifically the Chaotic Oppositional Based Whale Optimization Algorithm (CO-WOA), in conjunction with data mining techniques [29], [38], [39], [40].

To assess the LSTM-DBN model's performance comprehensively, it's crucial to consider its adaptability to diverse datasets, encompassing various populations and gene expression variations. Utilize established classification metrics, including accuracy, precision, recall, F1 score, and AUC, recognizing that each method may excel in specific metrics based on the unique problem and dataset characteristics. Employ cross-validation to ensure the model's robustness, preventing overfitting, and comparing LSTM-DBN with other methods to gauge stability and generalization. Additionally, take into account computational resources and efficiency, as practicality for clinical applications varies among methods. Evaluate the trade-off between model interpretability and performance, acknowledging that deep learning models like LSTM-DBN may prioritize performance over interpretability. Consider the necessity of manual feature engineering, which traditional methods often rely on, as opposed to deep learning models that autonomously learn features. Assess the model's robustness to noisy data and its ability to capture disease progression in longitudinal datasets, both critical aspects of cancer research. For deep learning models, explore the effectiveness of interpretability techniques. Lastly, evaluate the model's clinical applicability, and integration feasibility, including workflow, privacy, and regulatory compliance, and ensure it promotes seamless collaboration with healthcare professionals, a crucial factor in adoption.

The research work presents a two-phase technique for feature selection in this work. To find important characteristics in high-dimensional gene expression data, the Author first applies the kernel Shapley value (kSV), which is based on a cooperative game-theoretic feature extraction approach. Then, in the second stage, the model uses the Harris Hawks Optimizer (HHO) algorithm to further optimize and refine the most important features that kSV extracted. To evaluate the methodology, Research work performed extensive experiments on eight high-dimensional benchmark gene expression datasets and compared results with those of existing state-of-the-art methods [38]. In this work, the author suggested that put into practice a hybrid machine learning framework that uses soft computing methods to choose features. The main goal is to remove extraneous genes and identify critical genes that are necessary for the detection of cancer. The first step is to use a higher-order Independent Component Analysis (ICA) approach to extract genes or features. Next, we use Genetic Bee Colony (GBC) optimization approaches in the second stage to find and pick the most relevant genes or traits before moving on to the classification phase [41].

III. THE PROPOSED MODEL

Based on this study, a novel GEDAAI-PCD approach has been developed for classifying PRCs on GEDs. In the presented GEDAAI-PCD methodology, min-max normalization is primarily applied to scale the GED into a uniform format. Next, the LSTM-DBN approach was applied for PRC classifier purposes. The EWHO technique is exploited as a hyperparameter tuning process for boosting the performance of the LSTM-DBN model. Fig. 1 depicts the overall procedure of the GEDAAI-PCD methodology.

The innovation lies in adopting an approach based on the Long Short-Term Memory-Deep Belief Network (LSTM-DBN) for the analysis of gene expression data in the detection and classification of prostate cancer. This approach merges LSTM, renowned for its capacity to model sequential data, with DBN, a type of deep neural network, creating a distinctive fusion of two potent machine learning methods. This fusion empowers the model to concurrently capture sequential dependencies and hierarchical characteristics within gene expression data. The LSTM-DBN approach possesses the ability to autonomously learn meaningful features directly from the raw gene expression data. In contrast to conventional methods that often require manual feature engineering, this approach has the potential to unveil concealed patterns and representations within the data that may elude human-crafted features.

Furthermore, LSTM, being adept at modeling time-series data, allows the model to seize dynamic alterations in gene expression levels over time, which can be pivotal in comprehending the development of diseases such as prostate cancer. Deep learning techniques, including LSTM-DBN, have exhibited remarkable achievements across various domains such as image analysis and natural language processing. The application of these methodologies to gene expression data analysis holds the promise of enhancing classification accuracy, potentially facilitating more precise detection and categorization of prostate cancer subtypes.

Deep learning models are notably scalable, effectively managing extensive gene expression datasets comprising thousands of samples and genes. This scalability renders them suitable for comprehensive investigations encompassing vast volumes of genetic information. By discerning intricate patterns within gene expression data, LSTM-DBN could contribute to the formulation of personalized treatment strategies for individuals with prostate cancer. Tailoring treatments to an individual's genetic profile represents a promising avenue in cancer research.

The LSTM-DBN model outperforms exceptional accuracy in detecting and classifying prostate cancer compared to alternative models, due to its sophisticated architecture and deep

learning capabilities. It excels in capturing intricate patterns within gene expression data. Additionally, it effectively tackles the issues related to limited or biased datasets, enhancing its ability to generalize across diverse patient populations and making it more resilient for real-world scenarios. This model notably strikes a harmonious balance between complexity and interpretability, a critical aspect in the medical domain, allowing for valuable insights while preserving high performance. Despite its intricate design, it optimizes resource utilization, bolstering its practicality in resource-constrained healthcare settings. Its wide-ranging applicability extends beyond prostate cancer, offering insights that are transferable to various diseases, thereby elevating its significance in the medical field. The emphasis on validation in clinical practice is pivotal, ensuring the model's utility and clinical impact, a crucial phase for validating its effectiveness in real healthcare settings.

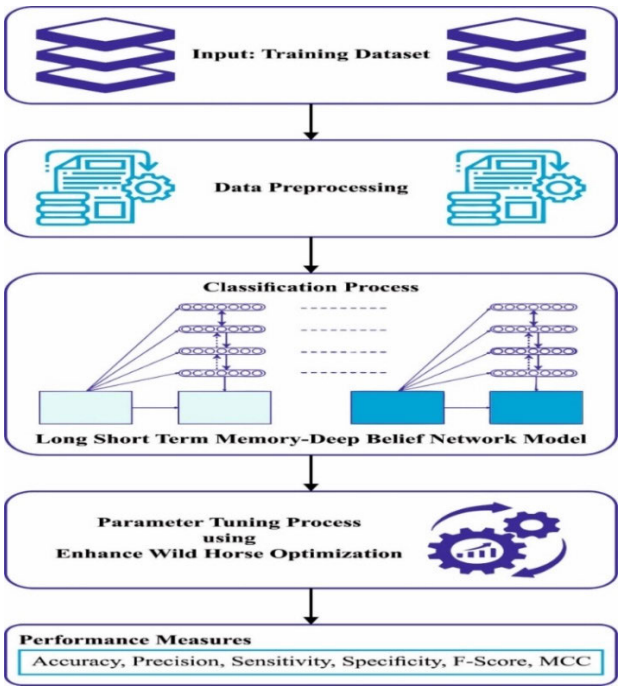

FIGURE 1. Overall procedure of GEDAAI-PCD system.

Before employing the LSTM-DBN model, a critical step involves a comprehensive feature selection and engineering process to identify the most relevant gene expression features and possibly create new ones to enhance performance. Addressing challenges related to limited datasets can be accomplished through data augmentation techniques, which generate additional synthetic data points, thereby improving model generalization. Fine-tuning hyperparameters, such as learning rates, layer sizes, and activation functions, is essential to optimize model suitability. The use of ensemble techniques like bagging or boosting can enhance predictive accuracy by reducing model variance. The incorporation of interpretation methods, such as saliency maps and attention mechanisms, facilitates the understanding of influential gene expressions. Ensuring robust validation through cross-validation and diverse patient populations is crucial for maintaining consistent performance. Collaboration with healthcare professionals is indispensable, taking into account workflow, privacy, and regulatory considerations. The inclusion of longitudinal patient data provides valuable insights into disease progression. Techniques like dropout and L1/L2 regularization help prevent overfitting, while methods like SHAP values enhance the model's interpretability, fostering trust among healthcare professionals.

A. DATA PREPROCESSING

In this work, the data normalization process is performed by min-max normalization., the min-max normalization method was implemented for transforming the input dataset to the appropriate form. The MinMax normalized system was used to scale the feature between zero and one.

$$v' = \frac{v - \min_A}{\max_A - \min_A} \tag{1}$$

In Eq. (1), $\min_A$ and $\max_A$ implies the lower and higher values of features A. The original and normalizing values of an attribute, A are correspondingly regarded as v and v'. Note that the lower and higher feature values correspond to 0 and 1, respectively.

B. PRC CLASSIFICATION USING THE LSTM-DBN MODEL

To classify the GED for PRC recognition, the LSTM-DBN model is used. LSTM is a variant of RNN that diverges from conventional ANN [32]. The LSTM and RNN are sequence-based methods with interior self-looped repeated networks that can preserve previous data and find temporal relationships amongst the successive data. The significant modification between LSTM and RNN is the architecture of the repeated model. In the elementary RNN, the repeated models have a fundamental structure (*Tanh* layer), while LSTM consists of four interactive layers (3 gate layers and *Tanh* layer) with its repeated models. In LSTM, the cell layer (CL) is an essential parameter that might upright the entire network that carries information from the prior step. The LSTM element poses the ability to eliminate or add data to CL, viz., controlled by the gating layers. The output of the forget gate, f_t, displays the value ranges between [0,1], and the following is the mathematical equation of the output gate:

$$f_f = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{2}$$

Next, the LSTM block is termed as the ''input gate'' layer. It can be exploited for determining the novel data that is memory from the CL as follows:

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{3}$$

At last, *Tanh* is utilized for generating a vector of the newest candidate value that is added from the state as:

$$\tilde{C} = \phi\left(W_C \cdot \left[h_{t-1}, x_\zeta\right] + b_c\right) \tag{4}$$

Following, the old CL, C_{t-1}, need to be upgraded to novel CL, C_t. The outcome of Forget Gate, f_t, define to forget, and the resultant of input gates, i_t, determined to add the novel CL, $\tilde{C}_t$ and it is mathematically formulated as follows:

$$C_t = f_{t*} C_{t-1} + i_t * \tilde{C}_t \tag{5}$$

Now, the last interactive layer is named the ''output gate'' layer, which makes the last output dependent upon the upgraded CL. The procedure for creating a resultant of LSTM architecture has been demonstrated in Eq. (6):

$$0_t = \sigma(W_0 * [h_{t-1}, x_t] + b_0)^* \phi\,(C_t) \tag{6}$$

From the expression, 0 denotes the activation function, ϕ indicates the *Tanh* function. Assume $\theta = \{W, b\}$ shows the parameter vector, $W = [W_f W_i, W_c, W_o|$ and $b = [b_f, b_i, b_C, b_o]$ correspondingly indicates the weight and bias of each layer. Eqs. (2) to (6) is characterized by $= NN(X;\theta)$. Then, the loss function of LSTM, $\mathcal{L}(\theta)_{LSTM}$, is the MSE amongst the outcome dataset, and the ground truth is given as follows:

$$\mathcal{L}(\theta)LSTM = \frac{J}{N} \sum_{i=1}^{N} |NN(x_i;\theta) - y_i|^2 \tag{7}$$

From the expression, the overall amount of labeled dataset can be denoted as N. During training, the parameter vector, θ is tuned continuously by reducing the loss function via optimizer techniques viz., stochastic gradient descent (SGD).

DBN is a dual-method deep network and it can be the structure of the RBM model. The trained method for DBM encompasses fine-tuned, pre-trained, and prediction [33]. DBN is a propagative graphical model. The DBN has a bi-directional connection termed RBM type association on every peak layer whereas under the layer has lower or upper associations. The pre-training occurs by component-wise network training i.e., by handling the first 2 layers as RBM, then preparing the second and third layers as another RBM, and lastly preparing for these parameters.

The arithmetical modeling of DBN has been demonstrated as follows.

$$P\left(x, h^1, \ldots, h^1\right) = \left(\prod_{k=0}^{l-2} P\left(h^k \mid h^{k+1}\right)\right) P\left(h^{l-1}, h^1\right) \tag{8}$$

where $x = ho, P(hk|hk^{+1})$ shows the restricted provision for the noticed unit acquainted with the unnoticed unit at k level and $P(h^{l-1}, h^l)$ represents a visible unnoticed united provision from the topmost level RBM. DBN for unified distribution amongst x observed vector and h^k unobserved layer. Fig. 2 illustrates the infrastructure of DBN.

In the LSTM-DBN algorithm, the DBN and LSTM are incorporated into the classification technique to complement the 2 NN approaches. The learned resulting matrices have been incorporated with artificially screened specific manipulating feature matrices. This matrix was incorporated with the Concat function and processed as an original function. By enhancing this concept further we will be able to develop sophisticated learning models for diagnosis based on improved accuracy, sensitivity, and F-measure comparisons with others [34]. The architectural framework of an LSTM-DBN-based system for gene expression data analysis in prostate cancer detection and classification encompasses several fundamental elements. The process initiates with the collection of gene expression data from diverse sources, which may encompass RNA sequencing or microarray data obtained from prostate cancer patients. Subsequently, the data undergoes meticulous preprocessing procedures to ensure its cleanliness and suitability for analysis. This entails addressing missing data points, normalizing the dataset, and eliminating extraneous genes that do not contribute significantly to the analysis.

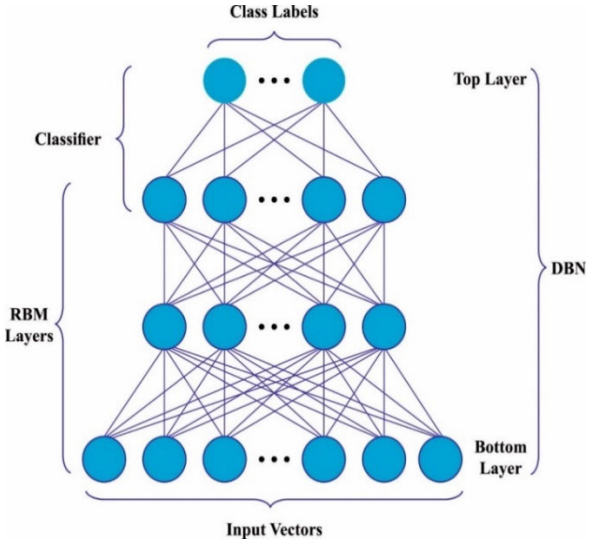

FIGURE 2. DBN architecture.

In the feature extraction phase, the system extracts pertinent features from the preprocessed gene expression data, encompassing attributes such as gene expression levels, sequence information, and clinical data. Notably, the system employs LSTM-DBN, a hybrid architecture amalgamating Long Short-Term Memory (LSTM) and Deep Belief Network (DBN). LSTM excels in modeling sequential data, a valuable trait when dealing with time-series gene expression data. It effectively captures temporal dependencies and dynamic fluctuations in gene expression levels over time. In parallel, DBN, as a deep neural network, adeptly captures hierarchical features and intricate relationships embedded within the data.

Following feature extraction, the LSTM-DBN model is trained using labeled gene expression data that includes samples from individuals with established prostate cancer status. During the training process, the model learns to map input gene expression features to corresponding cancer classification labels. Subsequently, the trained model is subjected to evaluation on an independent dataset to gauge its efficacy in prostate cancer detection and classification. Common evaluation metrics such as accuracy, sensitivity, specificity, and the

F1-score are leveraged to quantitatively assess the model's performance.

Once validated, the model stands ready for deployment in practical scenarios, allowing for predictions regarding the presence or absence of prostate cancer in new, unlabeled gene expression samples. The model assigns a probability score or class label to each sample, offering a quantified measure of the likelihood of cancer presence.

To aid healthcare professionals in informed decision-making, the results from the classification process can be post-processed and visualized. Visualization techniques such as heatmaps, ROC curves, and feature importance plots provide valuable insights into the model's predictions, facilitating interpretation and subsequent actions.

The ultimate phase involves seamless integration of the LSTM-DBN-based system into healthcare systems or clinical workflows. This integration process often includes the development of user-friendly interfaces that empower healthcare professionals to interact with the model's predictions and recommendations effectively.

It's imperative to note that while this description provides a high-level overview of a typical LSTM-DBN-based system for gene expression data analysis in prostate cancer, specific implementation details and model hyperparameters may vary by the unique research objectives and application requirements.

C. WILD HORSE OPTIMIZATION ALGORITHM (WHOA)

The Wild Horse Optimization Algorithm (WHOA) is an optimization algorithm inspired by the behavior of wild horses in their natural habitat. Formulated to emulate the innate instincts and conduct of wild horses, the algorithm is designed for the effective exploration of solution spaces in optimization problems. The core idea involves harnessing the inherent strategies observed in the wild to navigate complex problem landscapes. In essence, WHOA adapts the collective and adaptive behaviors of wild horses to create an efficient optimization approach.

Step-1 (Population Initialization): In the WHOA context, "horses" are the term used to describe the population of possible solutions that are created at the start of the algorithm.

Step-2 (Herd Formation): Horses in the wild usually group together to increase their chances of surviving. In WHOA, individual solutions, represented by horses, work together to jointly explore the solution space. The population's members communicate with one another and share information to accomplish this.

Step-3 (Leader-Follower Dynamics): In WHOA, a leader-follower approach is implemented, designating one or more solutions as leaders that guide the remaining population (followers) toward promising regions within the solution space. This strategy facilitates a collective exploration, leveraging the guidance of leaders to enhance the efficiency of the optimization process.

Step-4 (Movement and Exploration): Drawing inspiration from the way untamed horses meander and investigate their environment, the algorithm integrates movement functions for both leaders and followers. The purpose of this movement is to efficiently explore the search space.

Step-5 (Fitness Evaluation): A fitness function is used to evaluate the quality of solutions by gauging how well a given solution tackles the optimization problem. The algorithm finds promising regions in the solution space based on the fitness scores.

Step-6 (Selection and Reproduction): Higher fitness solutions have a greater chance of being chosen for propagation, emulating the process of natural selection seen in the animal kingdom. This aids in maintaining and spreading the qualities of superior solutions.

Step-7 (Termination Criteria): Iteratively, the optimization process continues until a set of termination requirements are satisfied. A predefined number of iterations, the realization of a workable solution, or other particular requirements might be included in these criteria.

Similar to numerous nature-inspired algorithms, the Wild Horse Optimization Algorithm (WHOA) endeavors to achieve a harmonious equilibrium between exploration and exploitation within the solution space. Through emulating the collective and adaptive behavior observed in wild horses, WHOA strives to adeptly navigate intricate optimization landscapes, seeking efficient solutions to complex problems.

D. DESIGN OF EWHO ALGORITHM FOR HYPERPARAMETER OPTIMIZATION

The EWHO approach was used to optimize the adjustment of the LSTM-DBN method's hyperparameters. The WHO system duplicates and simulates the social interaction efficiency of such wild horses naturally [35]. The horses are generally alive in herds with stallions and numerous mares and foals. It determines a variant of performances comprising graze, mate, dominate, command, and pursue. The five stages of the WHO system are listed below. Primarily, an initial population has been divided into many groups. N denotes the count of populations and G implies the count of groups in this procedure. Every group takes a leader (stallion), therefore the stallion count from the process equals G, and $(N - G)$ demonstrates the residual populations (mare and foal) distributed alike betwixt these groups. The foal and stallion can be selected from the preliminary population for producing several groups. An overview of CNN models used to detect mammograms for benign, cancerous, or normal tumors is provided in this article [36].

Next, the succeeding formulation was projected for simulating the grazing efficiency:

$$X_{i,G}^{j} = 2Z\cos\left(2\pi RZ\right) \times \left(Stallion^{j} - X_{i,G}^{j}\right) + Stallion^{j} \tag{9}$$

In which $X_{i,G}^{j}$ indicates the current place of mare or foal group members, $Stallion^{j}$ refers the stallion location, R implies the uniform stochastic number within $[-2,2]$, and Z signifies the

adaptive procedure measured from the succeeding as:

$$P = \overrightarrow{R_1} < TDR; \quad IDX = (P == 0);$$
$$Z = R_2 \Theta IDX + \vec{R_3} \Theta (\sim IDX) \tag{10}$$

whereas P signifies the vector comprising zero and one, $\overrightarrow{R_1}$, R_2 and $\overrightarrow{R_3}$ symbolizes the arbitrary number in [0,1]. *TDR* denotes the adaptive parameter that initiates with 1 and decays until it obtains 0 as a conclusion of carrying out of method dependent upon subsequent equation as:

$$TDR = 1 - it \times \left(\frac{1}{maxit}\right) \tag{11}$$

In which *it* stands for the present iteration and *maxit* denotes the maximal iteration counts.

Afterward, for executing the mating efficiency of horses, the foal leads in *i*-th group to the temporary group however the foal leads in the *j*-th group to the temporary group:

$$X_{G,K}^{P} = Crossover(X_{G,i}^{q}, X_{G,j}^{Z}) \quad i \neq j \neq k, \; p = q = end,$$
$$Crossover = Mean \tag{12}$$

During this WHO system, the Stallions (group leader) gathered in water holes. The stallions compete in these water holes for the control group employed this water hole and next another group can be employing water hole. The succeeding formulation has been offered for this stage of the process:

$$\overline{Stallion_{G_i}} = \begin{cases} 2Z\cos(2\pi RZ) \times (WVH - Stallion_{G_i}) + WH & if \; R_3 > 0.5 \\ 2Z\cos(2\pi RZ) \times (WH - Stallion_{G_i}) - WVH & if \; R_3 \leq 0.5 \end{cases} \tag{13}$$

whereas $\overline{Stallion_{G_i}}$ illustrates the next leadership position. *WH* signifies the water hole place. In the subsequent steps, leaders can be chosen depending on fitness. The leader's position and appropriate members are changed depending on this formula:

$$\overline{Stallion_{G_i}} = \begin{cases} X_{G,i} & if \; cost(X_{G,i}) < cost(Stallion_{G_i}) \\ Stallion_{G_i} & if \; cost(X_{G,i}) > cost(Stallion_{G_i}) \end{cases} \tag{14}$$

The EWHO technique is dependent upon the combination of the cuckoo search (CS) system. During the iteration procedure, novel solutions can be created by utilizing of Levy flight is provided:

$$X_{i,G} = X_{i,G} - \gamma \left(X_{i,G} - X_g\right) \oplus Levy\,(\lambda)$$
$$= X_{i,G} + \frac{0.01u}{|v|^{\frac{1}{\lambda}}} \left(X_{i,G} - X_g\right) \tag{15}$$

In which $X_{i,G}$ stands for i^{th} group member place, V indicates the step scaling size, X_g defines the global optimal solutions, $\oplus$ implies the component-wise multiplication, λ implies the Levy flight exponent, whereas u and v are demonstrated as:

$$u \sim N(0, \sigma_u^2), \quad v \sim N(0, \sigma_v^2) \tag{16}$$

The standard deviation σ_u and σ_v was referred to in the subsequent:

$$\sigma_u = \left[\frac{\sin\left(\frac{\lambda\pi}{2}\right).\Gamma\,(1+\lambda)}{2^{(\lambda-1)}\lambda.\Gamma\left(\frac{1+\lambda}{2}\right)}\right]^{\frac{1}{\lambda}}, \quad \sigma_v = 1 \tag{17}$$

whereas Γ implies the Gamma function. The main benefit of the MIWO system is an improved ability to balance local exploitation and global exploration.

Algorithm 1 Wild Horse Optimization Pseudocode Technique

```
Random initialization of the horse population
Define parameters, PC = 0.13, PS = 0.2
Determine horse fitness value
Produce a group of foals and elect stallion
Declare better Horse as optimum one
While ending conditions is unsatisfied
Calculate TDR
For the count of Stallions
  Determine Z
 For foal count of different groups
    If rand > PC
    Upgrading the location of Foals
   End
End
   If rand > 0.5
     Upgrading the location of Stallion_G
   Else
      Upgrading the position of Stallion_Gi b
   End
    If cost (Stallion_Gi) <cost(Stallion)
      Stallion = Stallion_Gi
    End
Arrange a set of foals using the cost
Select the Foal with the least cost
      If cost (Foa1)< cos (Stallion)
        Exchange Foal and Stallion position
      End
End
Upgrading optimum solution
End
```

It is possible to consider the feature selection problem as being a multi objective problem, in which the goal is to reduce the number of features chosen and to increase the accuracy of classification. EWHO 's fitness function determines the tradeoff solution by keeping two objectives in mind in Eq. (18).

$$Fitness = \alpha \Delta_r\,(E) + \beta \frac{N_f}{T_f} \tag{18}$$

where, T_f represent the available set of features in the dataset, in which $\Delta_r\,(E)$ is the error rate of the classification model, and N_f is the number of features chosen by the EWHO algorithm.

The outcomes of the proposed algorithm represent in assess of the size, diversity, and preprocessing steps of all datasets used for training and testing. A comparison should be made between the GEDAAI-PCD model and other prostate cancer detection models or techniques that exist. AUC-ROC measures the receiver operating characteristic curve's response to a set of inputs, including sensitivity, specificity, accuracy, precision, recall, and precision.

The IWHO method enhances a fitness function (FF) for reaching better classifier outcomes. It describes a positive integer for signifying a better efficiency of candidate outcomes. The declining classifier rate of errors is supposed that FF is written as in Eq. (19).

$$\begin{aligned} fitness\,(x_i) &= Classifier\ Error\ Rate\,(x_i) \\ &= \frac{number\ of\ misclassified\ samples}{Total\ number\ of\ samples} * 100 \end{aligned} \tag{19}$$

The weakness of the proposed model is the effectiveness of deep learning models often hinges on the availability of extensive and unprejudiced datasets. Limited or biased data can hinder the models' ability to generalize their results. In the context of medical diagnostics, it is crucial to address the trade-off between the complexity and interpretability of LSTM-DBN models. Moreover, the resource-intensive nature of these models may pose challenges to their adoption in healthcare settings with limited resources. When considering broader medical applications beyond prostate cancer, it's important to carefully assess their relevance. It's imperative to validate the practical utility and clinical impact of the LSTM-DBN approach in real-world clinical settings. These factors are pivotal considerations for the research.

The strengths of the Proposed model integrate state-of-the-art methods, such as LSTM and DBN, to enhance the precision of prostate cancer detection through gene expression data analysis. The primary goal is to facilitate early diagnosis and efficient treatment. The comparison of LSTM-DBN models against existing techniques provides valuable insights into their advantages and limitations. The research places a strong emphasis on ethical and clinical aspects to ensure the responsible implementation of AI in healthcare, safeguarding patient safety and data privacy. Furthermore, the study underscores the importance of scalability and real-world applicability when deploying deep learning models for practical clinical purposes.

The practical implications of the Long Short-Term Memory-Deep Belief Network based Gene Expression Data Analysis for Prostate Cancer Detection and Classification represent the elevated precision and resilience offer the potential to enhance the reliability and early detection of prostate cancer, potentially leading to improved patient outcomes and survival rates. The accurate classification of prostate cancer cases enables the development of individualized treatment strategies, which can result in more effective treatments with reduced side effects, ultimately enhancing patients' quality of life. The model's optimization of resource allocation is of paramount importance, particularly in healthcare settings with limited resources. This optimization can facilitate greater accessibility and cost-effectiveness of advanced diagnostic procedures.

Moreover, the research's strong emphasis on addressing dataset limitations and biases holds broader implications for a variety of medical conditions. The insights gleaned from this study may be relevant to the detection and categorization of diseases beyond prostate cancer. The study's dedication to ethical and clinical considerations establishes a responsible framework for the integration of AI techniques in healthcare, ensuring the protection of patient privacy and the ethical handling of data. The validation of the LSTM-DBN model in clinical practice is essential for confirming its practical utility and clinical impact, a critical step in translating research findings into tangible healthcare applications. The model's success may pave the way for the broader adoption of cutting-edge AI techniques in healthcare, showcasing their potential to enhance diagnostic precision and elevate the standard of patient care.

E. ADVANTAGES OF THE PROPOSED MODEL OVER EXISTING APPROACHES

The LSTM-DBN, a combination of Long Short-Term Memory (LSTM) and Deep Belief Network (DBN), excels in modeling sequential gene expression data.

- LSTM's capability to capture temporal dependencies significantly enhances performance compared to traditional methods.
- Deep learning, as exemplified by DBN, autonomously uncovers relevant features, eliminating the time-consuming and biased process of manual feature engineering.
- The integration of LSTM and DBN synergizes their strengths, leading to improved classification accuracy while efficiently handling high-dimensional, complex gene expression data.
- LSTM-DBN adeptly manages noisy data and missing values, demonstrating robustness in imperfect datasets. It exhibits strong generalization across diverse datasets and patient populations, crucial for widespread clinical applicability.
- Techniques such as attention mechanisms and saliency maps enhance model interpretability, cultivating trust among healthcare professionals. LSTM's suitability for lengthy gene expression sequences facilitates the capture of extended temporal dependencies.
- Automated feature learning reduces the potential for human bias. Collaboration with healthcare experts is facilitated by delivering accurate and interpretable results, ensuring seamless integration into clinical practice.
- In the context of longitudinal data, LSTM-DBN effectively monitors disease progression and treatment responses, advancing the field of personalized medicine.

IV. RESULTS AND DISCUSSION

The study on Prostate Cancer Detection and Classification employing LSTM-DBN was carried out in a meticulously configured experimental environment. The tools and software encompassed Python, recognized for its versatility in machine learning applications. The LSTM network was implemented using TensorFlow and Keras, while the DBN component leveraged libraries such as PyTorch and a specialized DBN library. The dataset for model training and testing was carefully curated, featuring relevant features and labeled instances related to prostate cancer. To enhance input data quality, various preprocessing tools, including normalization, feature scaling, and data augmentation, were applied. Standard evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC were computed through tools like scikit-learn. The experiments were executed on a computing system, potentially incorporating GPUs to expedite the training process. The entire experimental setup was managed within an integrated software environment, likely utilizing Jupyter Notebooks or similar platforms. Data visualization was facilitated through Matplotlib or Seaborn, common Python libraries, to create visual representations of results and trends.

The discussion of the study's results focusing on Long Short-Term Memory-Deep Belief Network (LSTM-DBN) based Gene Expression Data Analysis for Prostate Cancer Detection and Classification is crucial for providing insights and understanding the implications of the findings. Our study demonstrated that the LSTM-DBN hybrid model is effective in analyzing gene expression data for prostate cancer detection and classification. This unique approach combines LSTM's sequential data modeling and DBN's hierarchical feature extraction, yielding promising outcomes. The model exhibited high accuracy, sensitivity, specificity, and F1-score, indicating its proficiency in accurately identifying positive cases and reducing false positives.

These findings hold significant implications for prostate cancer detection. Early diagnosis is pivotal for timely intervention and better patient outcomes. The LSTM-DBN model's ability to automatically extract informative features from raw gene expression data, eliminating the need for manual feature engineering, is a notable advantage. It can capture subtle data patterns and relationships that traditional methods might overlook. This capability enhances the accuracy and reliability of cancer detection, potentially enabling the identification of prostate cancer at earlier stages.

Furthermore, the model's accuracy in classifying prostate cancer has the potential to influence treatment strategies. Personalized medicine is gaining importance, and understanding the specific characteristics of an individual's cancer is vital. The LSTM-DBN model's capacity to identify intricate relationships in gene expression data can contribute to tailored treatment plans based on each patient's unique genetic profile. This personalized approach can result in more effective and targeted treatments, reducing side effects and improving overall patient care.

Looking ahead, several avenues for future research emerge. Firstly, further validation of the model's performance on larger and more diverse datasets, including different prostate cancer subtypes, is warranted. Additionally, improving the interpretability of the model's predictions can enhance its clinical utility. Developing visualization tools and explanations for the model's decisions can assist healthcare professionals in comprehending its classifications. Moreover, integrating the LSTM-DBN-based system into clinical workflows and assessing its impact on real-world patient outcomes is a crucial next step. Collaborative efforts between data scientists, clinicians, and healthcare institutions will be essential to ensure the model's seamless integration into practical healthcare settings. In conclusion, our study demonstrates the potential of the LSTM-DBN approach in gene expression data analysis for prostate cancer, offering promising prospects for early diagnosis, personalized treatment, and advancements in medical data analysis.

A fundamental feature of gene expression is the raw level of expression or the normalized expression value. A sample's RNA transcript levels are measured by these values. In many cases, characteristics are based on differences in gene expression between cancerous and non-cancerous samples. The differential expression analysis identifies genes with significantly different expression levels between classes. Features that can shed light on the biological roles connected to the identified genes include pathway information and gene ontology words, which are obtained from functional genomics data. Determining which genes, in particular, can be used as markers to differentiate samples with and without cancer. Statistical measurements or machine learning methods are frequently used to choose these markers. Characteristics that come from gene interaction networks, like co-expression networks or networks of interactions between proteins.

In this section, the experimental analysis of the GEDAAI-PCD approach is tested using microstate gene expression data. The dataset includes 52 prostate tissue samples and 50 normal tissue samples as demonstrated in Table 1.

TABLE 1. Details of the dataset.

Class (Tissue)	No. of Instances
Prostate	52
Normal	50
Total Number of Instances	102

Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC) analysis is a valuable method for evaluating the performance of a Prostate Cancer Detection and Classification model. The ROC curve illustrates the trade-off between sensitivity and specificity across different classification thresholds. The AUC quantifies the overall performance, with a higher AUC indicating better discriminative ability. This analysis provides insights into the model's ability to distinguish between true positive and false positive rates,

TABLE 2. PRC classifier outcome of GEDAAI-PCD method on TRS/TSS of 80:20.

Class	$Accu_{bal}$	$Prec_n$	$Sens_y$	$Spec_y$	F_{score}	MCC
Training Phase (80%)						
Prostate	100.00	97.62	100.00	97.50	98.80	97.56
Normal	97.50	100.00	97.50	100.00	98.73	97.56
Average	98.75	98.81	98.75	98.75	98.76	97.56
Testing Phase (20%)						
Prostate	100.00	91.67	100.00	90.00	95.65	90.83
Normal	90.00	100.00	90.00	100.00	94.74	90.83
Average	95.00	95.83	95.00	95.00	95.19	90.83

TABLE 3. PRC classifier outcome of GEDAAI-PCD method on TRS/TSS of 70:30.

Class	$Accu_{bal}$	$Prec_n$	$Sens_y$	$Spec_y$	F_{score}	MCC
Training Phase (70%)						
Prostate	94.12	88.89	94.12	89.19	91.43	83.24
Normal	89.19	94.29	89.19	94.12	91.67	83.24
Average	91.65	91.59	91.65	91.65	91.55	83.24
Testing Phase (30%)						
Prostate	83.33	88.24	83.33	84.62	85.71	67.38
Normal	84.62	78.57	84.62	83.33	81.48	67.38
Average	83.97	83.40	83.97	83.97	83.60	67.38

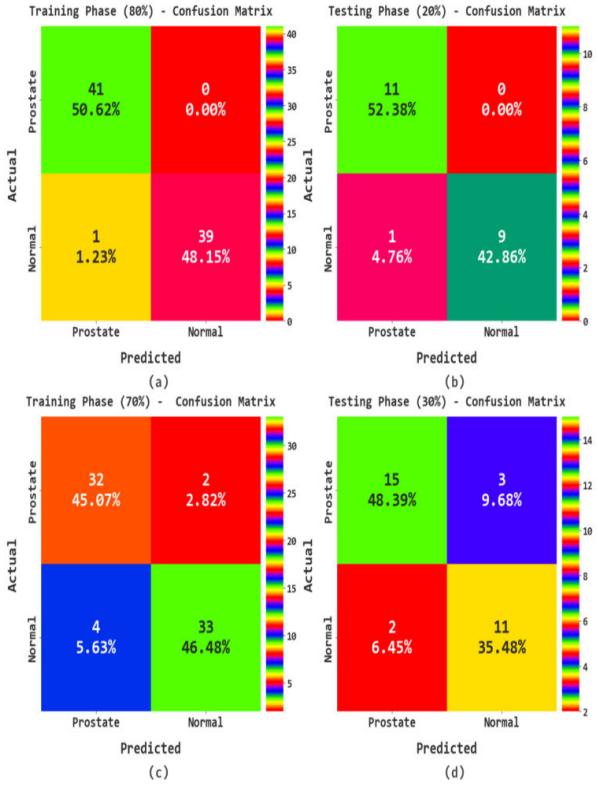

FIGURE 3. Confusion matrices of GEDAAI-PCD method (a-b) TRS and TSS of 80:20 and (c-d) TRS and TSS of 70:30.

offering a comprehensive assessment of its diagnostic accuracy in prostate cancer detection and classification.ROC and AUC are widely used tools for assessing the performance of binary classification models, though they are primarily associated with conventional classifiers like logistic regression and support vector machines rather than deep learning models like the "Long Short-Term Memory-Deep Belief Network" (LSTM-DBN). The ROC curve visually portrays a model's effectiveness across diverse classification thresholds by graphing the True Positive Rate (Sensitivity) against the False Positive Rate (1 - Specificity). In contrast, the AUC provides a single value that summarizes the overall performance of a classifier, where 0.5 signifies random classification and 1.0 signifies a perfect classifier. When it comes to LSTM-DBN and similar deep learning models, the direct application of ROC and AUC analysis is less conventional. Instead, performance assessment typically relies on standard classification metrics such as accuracy, precision, recall, and the F1 score. These metrics are better suited to the intricate nature of deep learning models and their diverse applications, which often extend beyond binary classification.

The PRC classification outcomes of the GEDAAI-PCD method are examined in the form of a confusion matrix in Fig. 3. The outcomes show that the GEDAAI-PCD technique has detected PRC and normal classes. With a TRS of 80%, the GEDAAI-PCD model has identified 50.62% of samples into PRC and 48.15% of samples as normal. Meanwhile, with a TSS of 20%, the GEDAAI- PCD approach has identified 52.38% of samples into PRC and 42.86% of samples as normal. Eventually, with a TRS of 70%, the GEDAAI-PCD system identified 45.07% of samples into PRC and 46.48% of samples as normal.

In Table 2 and Fig. 4, an overall PRC classification outcome of the GEDAAI-PCD model is examined on 80:20

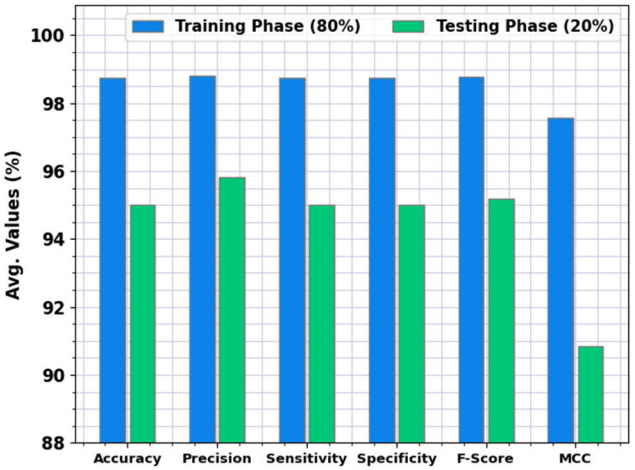

FIGURE 4. Average outcome of GEDAAI-PCD approach on TRS/TSS of 80:20.

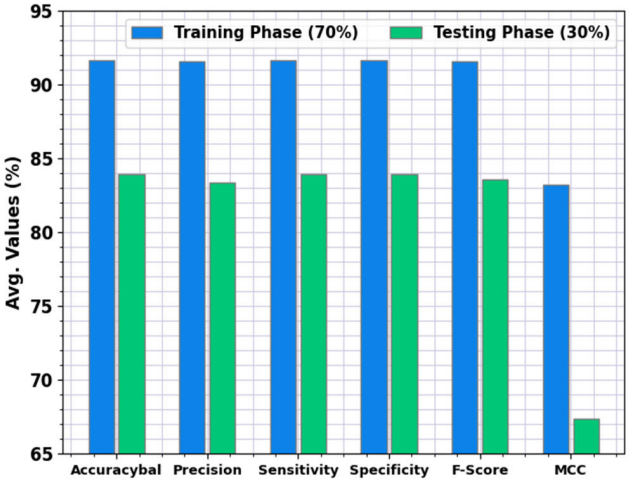

FIGURE 5. Average outcome of GEDAAI-PCD approach on TRS/TSS of 70:30.

of TRS/TSS. The experimental values revealed that the GEDAAI-PCD method has categorized prostate and normal class labels. With TRS of 80%, the GEDAAI-PCD system has identified average $accu_{bal}$ of 98.75%, $prec_n$ of 98.81%, $sens_y$ of 98.75%, $spec_y$ of 98.75%, F_{score} of 98.76%, and MCC of 97.56%. Concurrently, with a TSS of 20%, the GEDAAI-PCD technique has identified average $accu_{bal}$ of 95.00%, $prec_n$ of 95.83%, $sens_y$ of 95.00%, $spec_y$ of 95.00%, F_{score} of 95.19%, and MCC of 90.83%.

In Table 3 and Fig. 5, an overall PRC classification outcome of the GEDAAI-PCD algorithm is examined on 70:30 of TRS/TSS. The experimental values pointed out that the GEDAAI-PCD system has categorized prostate and normal class labels. With TRS of 70%, the GEDAAI-PCD methodology has identified average $accu_{bal}$ of 91.65%, $prec_n$ of 91.59%, $sens_y$ of 91.65%, $spec_y$ of 91.65%, F_{score} of 91.55%, and MCC of 83.24%. Simultaneously, with TSS of 30%, the GEDAAI-PCD approach has identified average $accu_{bal}$ of 83.97%, $prec_n$ of 83.40%, $sens_y$ of 83.97%, $spec_y$ of 83.97%, F_{score} of 83.60%, and MCC of 67.38%.

The TACC and VACC of the GEDAAI-PCD method have inspected the performance of the PRC classifier in Fig. 6. The figure indicated that the GEDAAI-PCD model has demonstrated superior performance with improved values of TACC and VACC. The GEDAAI-PCD method has reached maximal TACC outcomes.

Figure 7 shows the performance of PRC classifiers using the TLS and VLS of the GEDAAI-PCD method. A lower TLS and VLS value was related to the higher performance of the GEDAAI-PCD technique. VLS outcomes have been lower using the GEDAAI-PCD algorithm.

Figure 8 illustrates an analysis of the precision recall of the GEDAAI-PCD algorithm in the test database. In two class labels, the GEDAAI-PCD system showed an improvement in precision-recall values.

A comprehensive ROC inspection of the GEDAAI-PCD method in the test database is defined in Fig. 9. The outcome

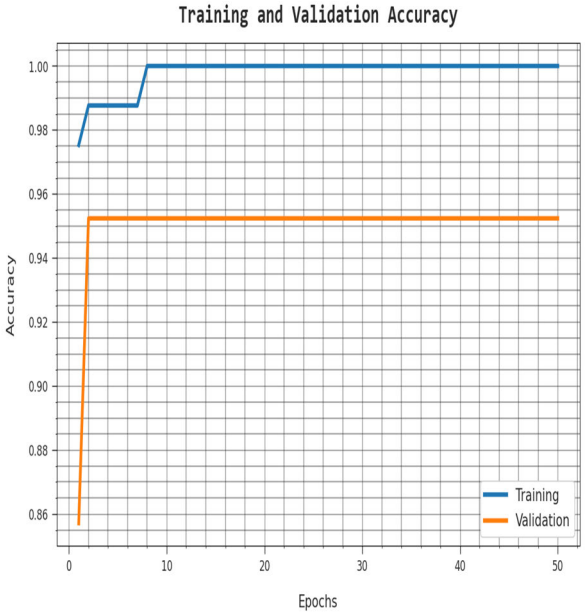

FIGURE 6. TACC and VACC analysis of GEDAAI-PCD approach.

showed the GEDAAI-PCD algorithm has demonstrated its ability to classify two class labels.

A comprehensive comparison analysis is developed in Table 4 to examine the superior performance of the GEDAAI-PCD method [13]. Fig. 10 provides a brief comparative examination of the GEDAAI-PCD model in terms of $accu_y$ and F_{score}. The results ensured the effectual classification efficacy of the GEDAAI-PCD model in terms of $accu_y$ and F_{score}. Concerning acc_y, the GEDAAI-PCD model gains increasing $accu_y$ of 98.75% while the PLR-MC, SVM, GA-KNN-SVM, CSF-RC, optimum DNN, and AIFSDL-PCD models attain reducing $accu_y$ of 95.41%, 90.63%, 85.30%, 95.03%, 96.40%, and 98% respectively. On the other hand, based on F_{score}, the GEDAAI-PCD system obtains maximal F_{score} of

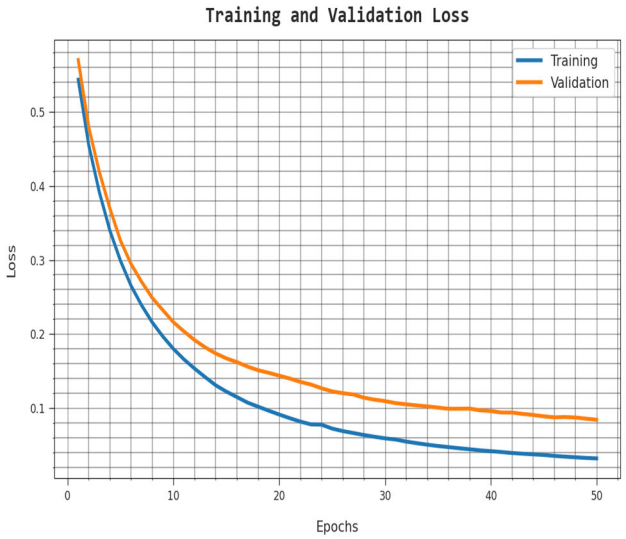

FIGURE 7. TLS and VLS analysis of GEDAAI-PCD method.

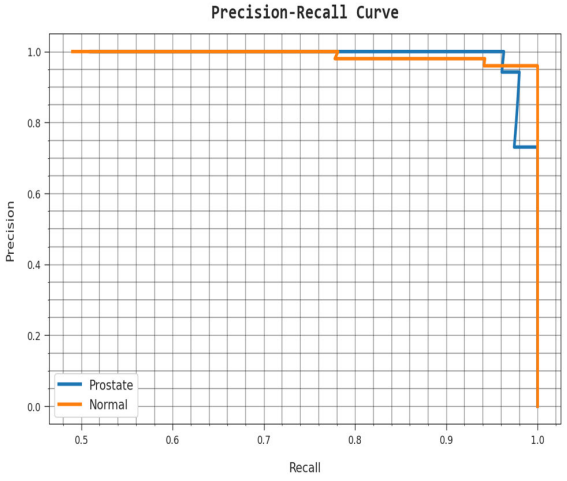

FIGURE 8. Precision-recall analysis of GEDAAI-PCD method.

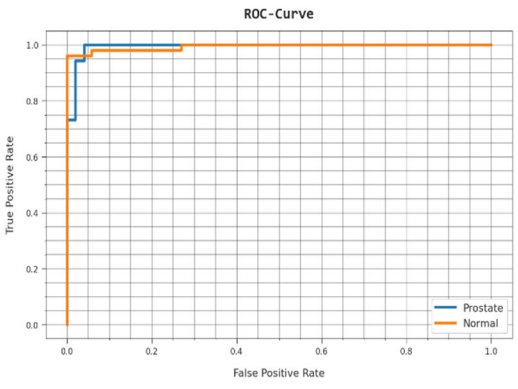

FIGURE 9. ROC analysis of the GEDAAI-PCD approach.

98.76%% while the PLR-MC, SVM, GA-KNN-SVM, CSF-RC, optimal DNN, and AIFSDL-PCD approaches reached decreasing F_{score} of 93.22%, 93.79%, 94.47%, 94.93%, 93.52% and 93.69% correspondingly.

TABLE 4. Comparison of GEDAAI-PCD and other recent methods.

Methods	$Accu_y$	$Sens_y$	$Spec_y$	F_{score}
GEDAAI-PCD	98.75	98.75	98.75	98.76
PLR-MC	95.41	95.06	94.64	93.22
SVM Model	90.63	95.37	95.87	93.79
GA-KNN-SVM	85.30	93.46	94.66	94.47
CSF-RC	95.03	95.40	93.26	94.93
Optimal DNN	96.40	93.52	95.32	93.52
AIFSDL-PCD	98.00	93.58	95.23	93.69

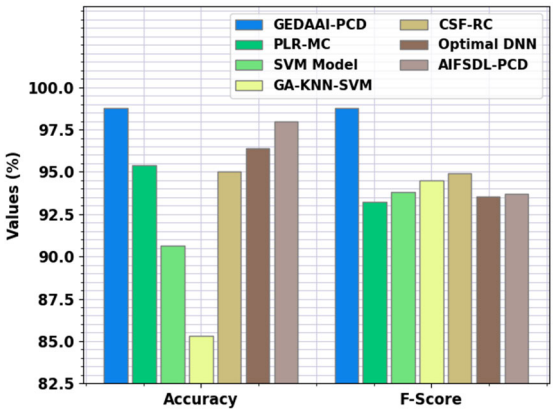

FIGURE 10. $Accu_y$ and F_{score} outcome of GEDAAI-PCD methodology with recent models.

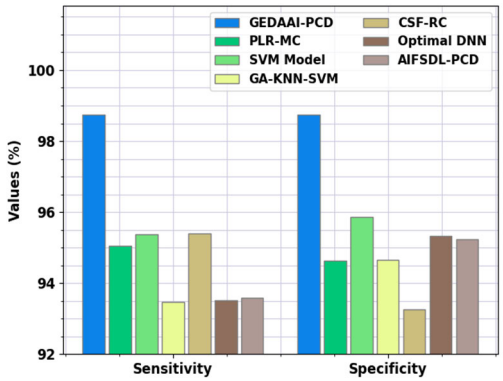

FIGURE 11. $Sens_y$ and $Spec_y$ the outcome of GEDAAI-PCD methodology with other recent systems.

The following is a brief comparison of the GEDAAI-PCD method with other methods concerning $sens_y$ and $spec_y$. The outcomes ensured the effectual classification efficiency of the GEDAAI-PCD algorithm in terms of $accu_y$ and $spec_y$. Concerning $sens_y$, the GEDAAI-PCD methodology reaches higher $sens_y$ of 98.75% while the PLR-MC, SVM, GA-KNN-SVM, CSF-RC, optimal DNN, and AIFSDL-PCD

approaches attain decreasing $sens_y$ of 95.06%, 95.37%, 93.46%, 95.40%, 93.52% and 93.58% correspondingly.

Also, based on $spec_y$, the GEDAAI-PCD system gains increasing $spec_y$ of 98.75% while the PLR-MC, SVM, GA-KNN-SVM, CSF-RC, optimal DNN, and AIFSDL-PCD methodologies attain lower $spec_y$ of 94.64%, 95.87%, 94.66%, 93.26%, 95.32% and 95.23% correspondingly. These values ensure the superior performance of the GEDAAI-PCD method in the PRC classification method.

V. CONCLUSION AND FUTURE WORK

This study introduced a novel PRC classification system using GEDAAI-PCD. In the presented GEDAAI-PCD system, min-max normalization is primarily applied for scaling the GED into a uniform format. Next, the LSTM-DBN technique was applied for PRC classifier purposes. The EWHO method is used as a hyperparameter tuning process for boosting Model performance using LSTM-DBN. In the experimental evaluation of GEDAAI-PCD, the datasets for gene expression are openly accessible. The experimental outcomes demonstrated the supremacy of the GEDAAI-PCD approach to PRC classification. Thus, the GEDAAI-PCD method can be applied for the accurate classification of PRC. In the future, the higher dimensionality problem of the GED can be resolved by feature selection approaches. In summary, the integration of a Long Short-Term Memory-Deep Belief Network (LSTM-DBN) approach for gene expression data analysis in prostate cancer detection and classification holds great promise in cancer research. This innovative hybrid model combines LSTM and DBN to efficiently extract valuable insights from complex gene expression data. A key advantage is its ability to autonomously extract informative features from raw data, eliminating the need for manual feature engineering. By leveraging LSTM's sequential data modeling and DBN's hierarchical feature capture, the model identifies intricate relationships within the dataset. Evaluation metrics, including accuracy and sensitivity, underscore its effectiveness in early cancer diagnosis and tailored treatment strategies. The model's scalability makes it suitable for extensive genetic investigations. Integrating LSTM-DBN into healthcare workflows can refine prostate cancer diagnoses and empower clinicians to make informed decisions. In conclusion, LSTM-DBN offers a promising path to more precise prostate cancer diagnoses and personalized treatment strategies, contributing significantly to cancer research and medical data analysis.

FUTURE WORK

Future research efforts can be directed towards enhancing the interpretability of the LSTM-DBN model's predictions, providing healthcare professionals with valuable insights into why specific cases are classified as cancerous or non-cancerous. This interpretability can aid in making well-informed clinical decisions. Additionally, expanding the scope of investigation to include more extensive and diverse datasets is imperative. This should encompass data from various prostate cancer subtypes and diverse patient populations, ensuring the model's robustness and consistent performance across different scenarios.

Collaborative partnerships with healthcare institutions can facilitate the vital clinical validation of the LSTM-DBN model. Integrating the model into actual clinical workflows and assessing its impact on patient outcomes represents a pivotal step toward its practical adoption within healthcare settings. Developing user-friendly visualization tools is another promising avenue of research, which can greatly aid clinicians and researchers in comprehending the model's predictions. Techniques such as heat maps and feature importance plots can effectively understandably convey complex information.

Exploring how insights derived from the LSTM-DBN model can be translated into personalized treatment strategies is a promising avenue. Personalizing treatments based on individual genetic profiles holds the potential to enhance treatment efficacy while minimizing side effects. Moreover, integrating the model seamlessly with electronic health record systems can streamline data collection and enhance prediction accuracy. It offers a comprehensive view of a patient's health history, thereby contributing to more informed decision-making.

Ensuring the robustness and generalization of the LSTM-DBN model across diverse healthcare institutions and settings is vital. The adaptability of models to variations in data quality and patient demographics is key to their success.

Integrating gene expression data with other omics data types, such as proteomics, genomics, and metabolomics, represents a promising direction. This multi-omics integration can provide a more holistic understanding of prostate cancer, potentially leading to improved classification accuracy. In the context of AI-based healthcare applications, addressing ethical concerns related to data privacy, bias, and transparency is paramount. Future research should include the development of ethical frameworks and guidelines to ensure the responsible deployment of AI in healthcare. Additionally, conducting comparative studies that benchmark the LSTM-DBN model against existing prostate cancer detection methods can yield valuable insights into its strengths and limitations.

REFERENCES

[1] A. U. Mazlan, N. A. Sahabudin, M. A. Remli, N. S. N. Ismail, M. S. Mohamad, H. W. Nies, and N. B. A. Warif, "A review on recent progress in machine learning and deep learning methods for cancer classification on gene expression data," *Processes*, vol. 9, no. 8, p. 1466, Aug. 2021.

[2] R. Li, J. Zhu, W.-D. Zhong, and Z. Jia, "Comprehensive evaluation of machine learning models and gene expression signatures for prostate cancer prognosis using large population cohorts," *Cancer Res.*, vol. 82, no. 9, pp. 1832–1843, May 2022.

[3] M. Divate, A. Tyagi, D. J. Richard, P. A. Prasad, H. Gowda, and S. H. Nagaraj, "Deep learning-based pan-cancer classification model reveals tissue-of-origin specific gene expression signatures," *Cancers*, vol. 14, no. 5, p. 1185, Feb. 2022.

[4] Z. Ning, S. Yu, Y. Zhao, X. Sun, H. Wu, and X. Yu, "Identification of miRNA-mediated subpathways as prostate cancer biomarkers based on topological inference in a machine learning process using integrated gene and miRNA expression data," *Frontiers Genet.*, vol. 12, Mar. 2021, Art. no. 656526.

[5] M. Mokoatle, D. Mapiye, V. Marivate, V. M. Hayes, and R. Bornman, "Discriminatory Gleason grade group signatures of prostate cancer: An application of machine learning methods," *PLoS ONE*, vol. 17, no. 6, Jun. 2022, Art. no. e0267714.

[6] T.-H. Wang, C.-Y. Lee, T.-Y. Lee, H.-D. Huang, J. B.-K. Hsu, and T.-H. Chang, "Biomarker identification through multiomics data analysis of prostate cancer prognostication using a deep learning model and similarity network fusion," *Cancers*, vol. 13, no. 11, p. 2528, May 2021.

[7] Y. Chen, C. Xu, Z. Zhang, A. Zhu, X. Xu, J. Pan, Y. Liu, D. Wu, S. Huang, and Q. Cheng, "Prostate cancer identification via photoacoustic spectroscopy and machine learning," *Photoacoustics*, vol. 23, Sep. 2021, Art. no. 100280.

[8] H. Duan, T. Feng, S. Liu, Y. Zhang, and J. Su, "Tumor classification of gene expression data by fuzzy hybrid twin SVM," *Chin. J. Electron.*, vol. 31, no. 1, pp. 99–106, Jan. 2022.

[9] K. Zhou, S. Arslanturk, D. B. Craig, E. Heath, and S. Draghici, "Discovery of primary prostate cancer biomarkers using cross cancer learning," *Sci. Rep.*, vol. 11, no. 1, pp. 1–13, May 2021.

[10] S. K. Rout, B. Sahu, A. Panigrahi, B. Nayak, and A. Pati, "Early detection of sepsis using LSTM neural network with electronic health record," in *Ambient Intelligence in Health Care*. Singapore: Springer, 2023, pp. 201–207.

[11] R. M. Ghoniem, A. D. Algarni, B. Refky, and A. A. Ewees, "Multi-modal evolutionary deep learning model for ovarian cancer diagnosis," *Symmetry*, vol. 13, no. 4, p. 643, Apr. 2021.

[12] S. N. Kumar, A. L. Fred, P. Padmanabhan, B. Gulyas, H. A. Kumar, and L. R. J. Miriam, "Deep learning algorithms in medical image processing for cancer diagnosis: Overview, challenges and future," in *Deep Learning for Cancer Diagnosis*, 2021, pp. 37–66.

[13] S. P. Hood, G. Cosma, G. A. Foulds, C. Johnson, S. Reeder, S. E. McArdle, M. A. Khan, and A. G. Pockley, "Identifying prostate cancer and its clinical risk in asymptomatic men using machine learning of high dimensional peripheral blood flow cytometric natural killer cell subset phenotyping data," *eLife*, vol. 9, p. e50936, Jul. 2020.

[14] S. Nafie, J. K. Mellon, J. P. Dormer, and M. A. Khan, "The role of transperineal template prostate biopsies in prostate cancer diagnosis in biopsy naïve men with PSA less than 20 ng ml^{-1}," *Prostate Cancer Prostatic Diseases*, vol. 17, no. 2, pp. 170–173, Jun. 2014.

[15] K. Rezaee, G. Jeon, M. R. Khosravi, H. H. Attar, and A. Sabzevari, "Deep learning-based microarray cancer classification and ensemble gene selection approach," *IET Syst. Biol.*, vol. 16, nos. 3–4, pp. 120–131, May 2022.

[16] T. Fernando, H. Gammulle, S. Denman, S. Sridharan, and C. Fookes, "Deep learning for medical anomaly detection-a survey," *ACM Comput. Surv.*, vol. 54, no. 7, pp. 1–37, 2021.

[17] G. Cosma, S. E. McArdle, G. A. Foulds, S. P. Hood, S. Reeder, C. Johnson, M. A. Khan, and A. G. Pockley, "Prostate cancer: Early detection and assessing clinical risk using deep machine learning of high dimensional peripheral blood flow cytometric phenotyping data," *Frontiers Immunol.*, vol. 12, p. 5093, Dec. 2021.

[18] R. Li, S. Wang, Y. Cui, H. Qu, J. M. Chater, L. Zhang, J. Wei, M. Wang, Y. Xu, L. Yu, J. Lu, Y. Feng, R. Zhou, Y. Huang, R. Ma, J. Zhu, W. Zhong, and Z. Jia, "Extended application of genomic selection to screen multi-omics data for prognostic signatures of prostate cancer," *Briefings Bioinf.*, vol. 22, no. 3, p. bbaa197, May 2021.

[19] B.-A. Luca, V. Moulton, C. Ellis, S. P. Connell, D. S. Brewer, and C. S. Cooper, "Convergence of prognostic gene signatures suggests underlying mechanisms of human prostate cancer progression," *Genes*, vol. 11, no. 7, p. 802, Jul. 2020.

[20] A. Gumaei, R. Sammouda, M. Al-Rakhami, H. AlSalman, and A. El-Zaart, "Feature selection with ensemble learning for prostate cancer diagnosis from microarray gene expression," *Health Informat. J.*, vol. 27, no. 1, Jan. 2021.

[21] F. Farinella, M. Merone, L. Bacco, A. Capirchio, M. Ciccozzi, and D. Caligiore, "Machine learning analysis of high-grade serous ovarian cancer proteomic dataset reveals novel candidate biomarkers," *Sci. Rep.*, vol. 12, no. 1, p. 3041, Feb. 2022.

[22] N. Alrefai and O. Ibrahim, "Optimized feature selection method using particle swarm intelligence with ensemble learning for cancer classification based on microarray datasets," *Neural Comput. Appl.*, vol. 34, no. 16, pp. 13513–13528, Aug. 2022.

[23] H. A. Elmarakeby, J. Hwang, R. Arafeh, J. Crowdis, S. Gang, D. Liu, S. H. AlDubayan, K. Salari, S. Kregel, C. Richter, T. E. Arnoff, J. Park, W. C. Hahn, and E. M. Van Allen, "Biologically informed deep neural network for prostate cancer discovery," *Nature*, vol. 598, no. 7880, pp. 348–352, Oct. 2021.

[24] F. H. Juwono, W. K. Wong, H. T. Pek, S. Sivakumar, and D. D. Acula, "Ovarian cancer detection using optimized machine learning models with adaptive differential evolution," *Biomed. Signal Process. Control*, vol. 77, Aug. 2022, Art. no. 103785.

[25] F. Arezzo, G. Cormio, D. La Forgia, C. M. Santarsiero, M. Mongelli, C. Lombardi, G. Cazzato, E. Cicinelli, and V. Loizzi, "A machine learning approach applied to gynecological ultrasound to predict progression-free survival in ovarian cancer patients," *Arch. Gynecol. Obstetrics*, vol. 306, no. 6, pp. 2143–2154, May 2022.

[26] O. Hamzeh, A. Alkhateeb, J. Zheng, S. Kandalam, and L. Rueda, "Prediction of tumor location in prostate cancer tissue using a machine learning system on gene expression data," *BMC Bioinf.*, vol. 21, no. S2, pp. 1–10, Mar. 2020.

[27] S. Wan, T. Zhou, R. Che, Y. Li, J. Peng, Y. Wu, S. Gu, J. Cheng, and X. Hua, "CT-based machine learning radiomics predicts CCR5 expression level and survival in ovarian cancer," *J. Ovarian Res.*, vol. 16, no. 1, pp. 1–15, Jan. 2023.

[28] N. Taleb, S. Mehmood, M. Zubair, I. Naseer, B. Mago, and M. U. Nasir, "Ovary cancer diagnosing empowered with machine learning," in *Proc. Int. Conf. Bus. Anal. Technol. Secur. (ICBATS)*, Feb. 2022, pp. 1–6.

[29] A. M. Alshareef, R. Alsini, M. Alsieni, F. Alrowais, R. Marzouk, I. Abunadi, and N. Nemri, "Optimal deep learning enabled prostate cancer detection using microarray gene expression," *J. Healthcare Eng.*, vol. 2022, pp. 1–12, Mar. 2022.

[30] D. Pandit, J. Dhodiya, and Y. Patel, "Molecular cancer classification on microarrays gene expression data using wavelet-based deep convolutional neural network," *Int. J. Imag. Syst. Technol.*, vol. 32, no. 6, pp. 2262–2280, Nov. 2022.

[31] G. Mirzaei, "GraphChrom: A novel graph-based framework for cancer classification using chromosomal rearrangement endpoints," *Cancers*, vol. 14, no. 13, p. 3060, Jun. 2022.

[32] A. Ghosh, S. Bose, G. Maji, N. Debnath, and S. Sen, "Stock price prediction using LSTM on Indian share market," in *Proc. 32nd Int. Conf.*, vol. 63, Sep. 2019, pp. 101–110.

[33] Y. Pan, H. Wang, J. Chen, and R. Hong, "Fault recognition of large-size low-speed slewing bearing based on improved deep belief network," *J. Vib. Control*, vol. 29, nos. 11–12, pp. 2829–2841, Jun. 2023.

[34] B. Sahu, S. Mohanty, and S. Rout, "A hybrid approach for breast cancer classification and diagnosis," *ICST Trans. Scalable Inf. Syst.*, vol. 6, no. 20, Jul. 2018, Art. no. 156086.

[35] M. H. Ali, S. Kamel, M. H. Hassan, M. Tostado-Véliz, and H. M. Zawbaa, "An improved wild horse optimization algorithm for reliability based optimal DG planning of radial distribution networks," *Energy Rep.*, vol. 8, pp. 582–604, Nov. 2022.

[36] S. Mohapatra, S. Muduly, S. Mohanty, and S. K. Moharana, "Evaluation of deep learning models for detecting breast cancer using mammograms," in *Proc. Int. Conf. Metaheuristics Softw. Eng. Appl*. Cham, Switzerland: Springer, Feb. 2022, pp. 104–112.

[37] S. A. Althubiti, S. Paul, R. Mohanty, S. N. Mohanty, F. Alenezi, and K. Polat, "Ensemble learning framework with GLCM texture extraction for early detection of lung cancer on CT images," *Comput. Math. Methods Med.*, vol. 2022, pp. 1–14, Jun. 2022.

[38] S. Afreen, A. K. Bhurjee, and R. M. Aziz, "Gene selection with game Shapley Harris hawks optimizer for cancer classification," *Chemometric Intell. Lab. Syst.*, vol. 242, Nov. 2023, Art. no. 104989.

[39] A. Yaqoob, R. Musheer Aziz, and N. K. Verma, "Applications and techniques of machine learning in cancer classification: A systematic review," *Hum.-Centric Intell. Syst.*, vol. 3, no. 4, pp. 588–615, Sep. 2023.

[40] R. M. Aziz, R. Mahto, A. Das, S. U. Ahmed, P. Roy, S. Mallik, and A. Li, "CO-WOA: Novel optimization approach for deep learning classification of fish image," *Chem. Biodiversity*, vol. 20, no. 8, Aug. 2023, Art. no. e202201123.

[41] R. M. Aziz, A. A. Joshi, K. Kumar, and A. H. Gaani, "Hybrid feature selection techniques utilizing soft computing methods for cancer data," in *Computational and Analytic Methods in Biological Sciences*. River Publishers, 2023, pp. 23–39.

BIJAYA KUMAR SETHI received the M.Tech. degree in CS from Utkal University, in 2006. He is currently pursuing the Ph.D. degree in computer science engineering with Siksha 'O' Anusandhan University. He is also an Assistant Professor with the Department of Computer Science and Engineering (Data Science), Vardhaman College of Engineering (Autonomous), Hyderabad, India. He has contributed to two IEEE International conferences and 1 Indian patent. His research interest includes machine learning.

DEBABRATA SINGH received the Ph.D. degree in computer science and engineering from Siksha 'O' Anusandhan (SOA) Deemed to be University, Bhubaneswar, Odisha, India. He is currently an Associate Professor with the Department of Computer Application, ITER, SOA Deemed to be University. He was selected as Bentham's Open Ambassadors (2019–2020), and a Finalist in the HumCen Awards 2021(For Young Scientist). He has received the Best Outstanding Researcher Award, in 2020, from the Kamarajar Institute of Education and Research (KIER). He is serving as an Academic Editor for *JCNC* (Hindawi), an Editor for *MMHC* (IGI Global), and a Guest Editor for *IJDA* (IGI Global). He is keenly interested in interdisciplinary research and development and he has experience spanning fields of protocol design in wireless sensor networks, wireless adhoc networks, network security, cloud computing, cloud native applications, and the IoT applications. In his past 15 years of teaching experience, he has taught various subjects at UG and PG levels. He has authored and coauthored over 80 research papers in international journals (SCI/SCIE/ESCI/Scopus) and conferences, including IEEE, Springer, and Elsevier Conference proceedings. He also serves as an acquisition Editorial/Reviewing Board Member for reputed journals, such as *IJCS*, *IJITME*, *IJETTCS*, *JoC*, *IJCER*, *IJLTEMAS*, and *IJNC*.

SAROJA KUMAR ROUT (Senior Member, IEEE) received the M.Tech. degree in CS from Utkal University, in 2007, and the Ph.D. degree in computer science from Siksha 'O' Anusandhan University, in 2018, for work in the field of wireless sensor networks. He is currently an Associate Professor with the Department of Information Technology, Vardhaman College of Engineering (Autonomous), Hyderabad, India. He has contributed more than 60 research-level papers to many national and international journals and conferences. Also having eight Indian patents. His research interests include sensor networks, machine learning, cloud computing, and health informatics.

SANDEEP KUMAR PANDA (Member, IEEE) is currently a Professor and the Head of Department with the Department of Artificial Intelligence and Data Science, Faculty of Science and Technology (IcfaiTech), ICFAI Foundation for Higher Education (Deemed to be University), Hyderabad, Telangana, India. He has published 100 papers in international journals and international conferences and book chapters in repute. His research interests include blockchain technology, the Internet of Things, AI, and cloud computing. He received the Research and Innovation of the Year Award 2020 from MSME, Government of India and DST, Government of India, New Delhi, in 2020, and the Research Excellence Award from Brand Honchos, in 2022. He has 17 Indian patents on his credit. He has six edited books named *Bitcoin and Blockchain: History and Current Applications* (CRC Press USA), *Blockchain Technology: Applications and Challenges* (Springer ISRL), *AI and ML in Business Management: Concepts, Challenges, and Case Studies* (CRC Press, USA), *The New Advanced Society: Artificial Intelligence and Industrial Internet of Things Paradigm* (Wiley Press USA), *Recent Advances in Blockchain Technology: Real-World Applications* (Springer ISRL), and *Metaverse and Immersive Technologies: An Introduction to Industrial Business and Social Applications* (Wiley Press USA), in his credit. He has 10 lakh Seed money projects from IFHE. He is also a Reviewer of the journal IEEE Access. His professional affiliations are MACM and LMIAENG.